ION INTERACTIONS WITH SINGLE- AND DOUBLE-STRANDED DNA MEASURED USING SOLUTION X-RAY SCATTERING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Stephen Paul Meisburger

August 2014

ION INTERACTIONS WITH SINGLE- AND DOUBLE-STRANDED DNA MEASURED USING SOLUTION X-RAY SCATTERING

Stephen Paul Meisburger, Ph.D.

Cornell University 2014

Salt ions and water present in living cells are essential mediators of protein-nucleic acid interactions. X-ray scattering and computational methods are applied to study the ion distributions around double-helical DNA, and to understand the complex electrostatic interactions that modulate single-stranded DNA conformation in solution. The spatial distributions and numbers of ions are measured and compared with theoretical predictions. X-ray data for single-stranded DNA reveal ion- and sequence-dependent properties of the conformational ensembles that support recent predictions of polyelectrolyte theory and provide a structural basis for the mechanical properties of these nucleic acids. The data presented here also identify shortcomings of current electrostatic theory and molecular dynamics force fields for modeling DNA in solution, and suggest how they may be improved. Finally, cryo-cooling methods are developed for x-ray scattering that provide unique advantages for high throughput structural investigations of biomolecules.

BIOGRAPHICAL SKETCH

Stephen (Steve) Meisburger was born in Poughkeepsie New York and grew up in California's San Francisco Bay Area. While a high school student at Bellarmine College Preparatory in San Jose, he was a member of FIRST robotics Team 254 and worked as a vacuum technician at the Stanford Synchrotron Radiation Laboratory. From 2003-2007, he attended Carleton College in Northfield Minnesota. He spent the summer of 2005 at Stanford University in a Research Experience for Undergraduates (REU) program run by the National Nanotechnology Infrastructure Network. There, he characterized aberrations in a uniform-field electron microscope in the lab of Fabian Pease. In 2006, he participated in the REU for Complex Systems at the Santa Fe Institute in New Mexico and worked on dynamical systems theory for mobile robot behaviors with Alfred Hübler. After receiving a B.A. in Physics in 2007, he began a Ph.D. program in Applied Physics at Cornell University in Ithaca, New York. He has worked in the lab of Lois Pollack since 2008.

Steve is an avid homebrewer, and his creations have earned a reputation for excellence: of one brew, called "Paper Maker", Dr. Drew Brisbin wrote, "Yo Steve, this coffee beer is the bomb! [...] I'm at the office working on my paper as I drink it." The recipe for "Paper Maker" is as follows: Infusion mash 11.5 lb Maris Otter, 1 lb English medium crystal, and 0.25 lb English chocolate malt with Ithaca city water for 60 minutes at 152 F. Sparge and collect 6.5 ga wort. Boil for 60 minutes: add 2 oz Willamette hops (60 min. remaining), 1 oz Willamette (10 min. remaining) and 1 lb D-180 Candi Syrup (flame-out). Chill, transfer to fermentor, pitch 1 packet Danstar Windsor ale yeast, aerate and ferment for 2 weeks. Transfer to secondary fermentor and add cold-brewed Gimme! coffee (12 oz Guatemala single origin, coarse grind, add water to 1 L total volume, steep 24 hours in refrigerator, filter). Ferment 2 more weeks, then bottle condition and share with friends.

to my family

ACKNOWLEDGEMENTS

First and foremost, I thank my advisor, Lois Pollack, for her patient support, optimism, encouragement, and wisdom. My research at Cornell would not have been possible without my talented colleagues in the Pollack group: Suzette Pabit, Li Li, Jessica Lamb, Josh Blose, Julie Sutton, Chris Jones, Andrea Katz, George Calvey, Yujie Chen, Josh Tokuda, and Huimin Chen. When I joined the lab, Suzette, Li, and Jessica taught me the basics of synchrotron X-ray scattering and nucleic acid biophysics, and I am very grateful for the skills they imparted and the example they set as outstanding experimentalists. Synchrotron X-ray experiments are necessarily a team effort, and I have had the pleasure of working closely with everyone mentioned above for in the twenty three synchrotron runs (give or take a few) since 2008. Thank you also to Huimin for guiding me through my first paper-writing experience, and for staying up all night to collect the X-ray data in Chapters 2-5. Pollack group members also helped with sample preparation: most of the ion-counting samples in Chapter 5 were prepared by Peter Gu as part of his undergraduate thesis in Applied and Engineering Physics, and Suzette Pabit prepared the duplex DNA used in Chapter 1.

I thank my special committee members, Sol Gruner and Ailong Ke, for their advice and critical reading of this manuscript, and Brian Crane for filling in at the last minute. I have benefited immeasurably from the accumulated knowledge of many previous students in the Pollack and Gruner groups; some I know only through their dissertations, well-documented code, websites, and published papers; and others I have been lucky enough to meet and collaborate with. The theory of ion contrast in Chapters 1 and 2 extends work by Xiangyun Qiu that he did while in the Pollack group. I owe special thanks to Kurt Andresen for hosting me at Gettysburg, collecting the atomic emission spectra in Chapter 5, and helping me prepare for my talk at the 2014 Biophysical Society Meeting.

Most of the experiments I describe were performed at the Cornell High Energy Synchrotron Source (CHESS). The CHESS and MacCHESS staff, especially Bill Miller, Scott Smith, and

scientists Arthur Woll, Ken Finkelstein, and Richard Gillilan were wonderful to work with and were always generous with expert and material assistance. Advances in X-ray instrumentation were essential for collecting high-quality data. The "sample oscillator" flow cell turned out to be a key innovation. The original design is due to Richard Gillilan, who generously allowed then-undergraduate Jiecheng (Betty) Zhang to copy and improve upon it. The version used in Chapters 1-6 includes custom software and hardware improvements by Suzette Pabit, Andrea Katz, Yujie Chen, and Josh Tokuda.

The all-atom molecular dynamics calculations in Chapters 2 and 3 were performed by Ron Elber and Serdal Kirmizialtin (U.T. Austin). Robert Thorne and Matt Warkentin (Cornell) initiated the cryo-SAXS project described in Chapter 6, inspired by Richard Gillilan's early experiments. Thorne group members including Jesse Hopkins contributed their considerable expertise with cryo-cooling and assisted with data collection. Chapter 1 would never have been written without a series of conversations with David Case and Hung Nguyen (Rutgers), and the inspiration for writing Chapter 3 came from conversations with Alan Chen (University at Albany-SUNY).

As a graduate student, I have had the privilege of working on many different projects that furthered my development as a scientist but did not reach publication stage. Collaborators Kenneth Hampel and Krista Brooks (University of Vermont) provided RNA samples, and later DNA templates and primers, for time-resolved folding experiments. I am grateful to Ailong Ke's lab, especially Fran Ding, for patiently teaching me some basic skills of molecular biology, such as PCR and gel electrophoresis, when I decided to make RNA myself. Josh Blose and Mark Landy helped to develop size exclusion purification protocols for Mark's undergraduate thesis in Applied and Engineering Physics. Micromixer fabrication was carried out at the Cornell NanoScale Science & Technology Facility (CNF), which is funded by the National Science Foundation. Thank you to CHESS, in particular Arthur Woll and Tom Szebenyi, for assisting with capillary focusing optics. Because of the generous support I received from these individ-

uals and facilities, I feel guilty for not finishing these projects in time to include them in the dissertation. However, I have left the projects in very capable hands: I look forward to seeing the mixer designs that George Calvey and Yujie Chen build, and hearing about Josh Tokuda's new RNA folding experiments. Andrea Katz and Jesse Hopkins have already taken the cryo-SAXS method beyond what I thought was possible, and I can't wait to read their papers.

Chapter 2 was previously published in Biopolymers, and is reprinted with permission from John Wiley and Sons. Chapter 6 was published in the Biophysical Journal and is reprinted with permission from Elsevier. I was supported financially by government funding agencies, primarily by the National Institutes of Health through the Molecular Biology Training Grant administered by Michelle Wang and Robert Oswald. I additionally received support from the National Science Foundation, and I am grateful for the support from the Cornell Laboratory for Accelerator-based Sciences and Education during my final year. Further acknowledgments, including specific grant numbers, are provided at the end of each Chapter.

Finally, I thank my family for their unwavering faith in my potential, and my Ithaca family; Drew, Betsey, Ben, and especially, Nora, for making these seven years a wonderful adventure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

QUANTITATIVE X-RAY SCATTERING MEASUREMENT OF IONS AND WATER AROUND DNA

Steve P. Meisburger[1] Suzette A. Pabit[1], and Lois Pollack[1]

1. Applied and Engineering Physics, Cornell University, Ithaca, New York

Abstract

Interactions between proteins and highly charged molecules such as DNA and RNA are essential for biological function. The energetics and specificity of these interactions depend on ions and water present near molecular surfaces. However, diffuse ion and water distributions have been difficult to characterize experimentally. Here we describe a small angle x-ray scattering method that provides sensitive and quantitative measurements of ion and water distributions around nucleic acids. Scattering profiles recorded on an absolute scale are analyzed using contrast variation techniques to separate the scattering from components. As proof of principle, we measure the spatial distribution of ions around double-stranded DNA in 100 mM monovalent salt. We find that counterions are present near the molecular surface in greater numbers than predicted by nonlinear Poisson-Boltzmann theory. Our results underscore the important role of water as a mediator of electrostatic interactions between DNA and its counterions: both ion and solvent degrees of freedom are important for accurate modeling of protein-DNA interactions in biological systems.

1.1 Introduction

In all known living organisms, genetic information is encoded in the structures of nucleic acids: DNA and RNA. These highly charged molecules attract an "atmosphere" of ions present in the cellular environment. When processing and transmitting the genetic code, cells use protein and RNA-based machines that interact directly with nucleic acids, often inducing structural transitions in order to access base-sequence information. In these essential protein-nucleic acid interactions, the ion atmospheres are necessarily altered. Theoretical calculations show that the change in electrostatic (or solvation) free energy accompanying such a transition or binding event may contribute significantly to the total free energy difference. In particular, measurements of equilibrium constants for RNA-protein interactions and RNA folding are often highly salt-dependent, in agreement with theoretical predictions [1, 2].

The electrostatic properties of RNA and DNA are also important for their collective behavior: DNA molecules strongly other in low-salt solutions [3], yet form condensed phases when certain multi-valent cations are added [4–6]. The ion atmosphere is also an important component of DNA-based nanotechnology. For example, specific counter-ion effects enable controlled crystallization of DNA-conjugated nanoparticles [7], and the ion atmospheres around surface-tethered molecules can be modified by electric fields for dynamically switching the hybridization state of DNA [8] and for label-free electronic detection [9–12].

The properties of the ion atmosphere have been understood from several theoretical perspectives. The counterion condensation (CC) theory for rod-like polyelectrolytes [13] divides the ions into energetically distinct "condensed" and "diffuse" layers. Theories based on the nonlinear Poisson-Boltzmann (NLPB) Equation describe all ions equivalently and apply more generally to different geometries [14]. All-atom grid-based NLPB solvers are now widely used for macromolecular electrostatics calculations [15–17]. NLPB and CC theories have several

simplifying assumptions that have been called into question for highly charged molecules such as DNA [14, 18–21]: ions are modeled as point-charges, ion-ion interactions are neglected, and the solvent is approximated as a uniform dielectric medium.

Early NMR studies of the ion atmosphere around DNA showed qualitative agreement with theory, but only counterions within a short and poorly-defined distance from the surface were detected [22]. On the other hand, small angle x-ray scattering (SAXS) is uniquely sensitive to the spatial distribution of all statistically-associated ions [19, 23, 24]. X-ray scattering signals contain contributions from the ions, macromolecule, and solvent that can be difficult to separate. Thus, several studies have taken advantage of the unique contrast mechanism of anomalous (or resonant) scattering [19, 25], where ion contrast is varied by tuning the x-ray energy near an absorption edge for the ion. Anomalous small angle x-ray scattering (ASAXS) provides a means for separating the ion-dependent component of the signal, and has been particularly useful for comparison with theory [19, 26, 27].

The thermodynamically important preferential interaction coefficient, equal to the number of associated ions divided by the macromolecule's charge, can be measured using the Donnan effect in equilibrium dialysis experiments [28, 29] or with ion-binding fluorescent dyes [30]. Recently, we introduced an ASAXS-based method for counting the total number of statistically associated ions around RNA and DNA [31]. X-ray scattering measurements are made on a calibrated absolute intensity scale (electrons^2 per molecule), so that the forward scattering amplitude depends linearly on the number of statistically-associated resonant counterions times the energy-dependent scattering factor of the ion. The ion scattering factors and the forward scattering amplitudes are measured at two or more x-ray energies, yielding the number of excess ions per molecule. Ion numbers measured for RNA and DNA agree with similar measurements by equilibrium dialysis and with theoretical predictions [26, 32, 33].

In ASAXS, the difference between scattering curves acquired at two energies provides in-

formation about the spatial distribution of ions. However, the achievable signal-to-noise ratio for ASAXS is limited by the small difference signals (typically less than 10%) and sensitivity of macromolecule solutions to radiation damage effects [34]. Furthermore, the macromolecule and its hydration layer contribute to the anomalous difference signal; these cross-terms must be taken into account when comparing with theory. Thus, to provide greater signal-to-noise, and to separate contributions from the nucleic acid and water, we extend the absolute intensity ASAXS method using the non-resonant contrast mechanism of heavy-ion substitution [19, 23, 24, 33, 35]. Relative to anomalous scattering, the variation in contrast with heavy-atom substitution can be large, providing more precise information about the locations of ions around the DNA. Absolute intensity calibration provides additional information, including the number of ions as well as the DNA hydration.

As a proof of principle measurement, we acquire heavy-atom and anomalous scattering data for DNA in 100 mM monovalent salt. By comparing predicted and measured SAXS profiles on an absolute scale, we show that the NLPB equation underestimates the number of ions present near the DNA surface. This result supports earlier conclusions from molecular dynamics that the dielectric environment near the nucleic acid surface is significantly different from the bulk [21]. These data should be valuable for testing new computational methods that build on NLPB theory by including solvent degrees of freedom [36, 37].

1.2 Background

In a dilute solution of molecules, the total X-ray scattering intensity is proportional to the contribution of a single molecule, averaged over all orientations and configurations (denoted by $\langle\rangle_{\Omega,\Pi}$):

$$I(q) = \left\langle |A(q)|^2 \right\rangle_{\Omega,\Pi}$$

where $A(q)$ is the (complex) scattering amplitude, equal to the Fourier transform of the electron density. When the electron density has two components, $A(q) = A_1(q) + A_2(q)$, the intensity involves the scattering from each component alone and a cross-term:

$$\begin{aligned} I(q) &= \left\langle |A_1(q)|^2 \right\rangle_{\Omega,\Pi} + 2\left\langle \mathrm{Re}\left(A_1(q)A_2^*(q)\right)\right\rangle_{\Omega,\Pi} + \left\langle |A_2(q)|^2 \right\rangle_{\Omega,\Pi} && (1.1) \\ &= I_1(q) + I_{12}(q) + I_2(q) && (1.2) \end{aligned}$$

Here, we examine a two-component system consisting of (1) the hydrated DNA molecule and (2) its counterion atmosphere. The goal of contrast variation is to separate the scattering contributions from these components. If the contrast of the molecule is δ_M, the contrast per ion is δ_I, and the number of excess ions is N_I, Equation 1.1 becomes

$$I(q) = \delta_M^2 P_M(q) + 2\delta_M(\delta_I N_I)P_{MI}(q) + (\delta_I N_I)^2 P_I(q) \tag{1.3}$$

The shape functions $P(q)$ are defined so that $P(0) = 1$. At $q = 0$, Equation 1.3 is factorable and the forward scattering amplitude $|A(0)| = I(0)^{1/2}$ is a linear function of the contrasts

$$I(0)^{1/2} = \delta_M + N_I\delta_I \tag{1.4}$$

When the intensity is measured in units of electron^2 per molecule, $I(0)^{1/2}$ equals the total number of scattering electrons for the molecule and ion components $Z_M + N_I Z_I$, minus the number of solvent electrons excluded:

$$\delta_M + N_I\delta_I = Z_M + N_I Z_I - \bar{\rho}_e V \tag{1.5}$$

where V is the change in system volume due to the presence of the DNA and its counterions (partial molar volume) and $\bar{\rho}_e$ is the average electron density of the solution. For hydrated ions, the effective number of scattering electrons per ion is

$$Z_{\text{eff}} = Z_I - \bar{\rho}_e V_I \tag{1.6}$$

where V_I is the change in system volume upon insertion of a single ion (partial molar volume). Note that V_I can be negative for small ions that strongly attract water (see Table 1.1). For the present analysis, we make the simplifying assumption that all of the ions are hydrated, which is supported by NMR measurements of group I cations around DNA [22]. Then, Equations 1.5 and 1.6 can be combined with $\delta_I = Z_{\text{eff}}$ to give the contrast of the DNA

$$\delta_M = Z_M - \bar{\rho}_e V_M \tag{1.7}$$

where V_M is the contribution to the partial molar volume from the DNA component ($V = V_M + V_I N_I$) that includes both hydration (increased water density near the surface) and excluded volume effects. In this work, the electron density of the solvent $\bar{\rho}_e$ is assumed to be that of liquid water, a good approximation for the conditions measured here (in the most non-ideal case of 100 mM CsCl, $\bar{\rho}_e$ differs from water by $\sim 1\%$). Using the fact that a water molecule has 10 electrons, Equation 1.7 can also be written

$$\delta_M = Z_{DNA} + 10\,(N_H - N_E) \tag{1.8}$$

where N_H and N_E are the numbers of hydrating and excluded waters per DNA molecule.

Experimentally, $I(q)$ is measured on an absolute scale for DNA in three or more salt solutions, where the ions have different x-ray contrasts but are chemically similar (e.g. group I monovalent ions). First, δ_M and N_I are found from a linear fit of Equation 1.4 to the extrapolated amplitudes, $I(0)^{1/2}$. Then, all of the coefficients in Equation 1.3 are known, and least-squares methods are applied to solve for the q-dependent terms (described in Section 1.5.5).

	Z	V_I (Å^3)	Z_{eff}	R_{M-O}(Å)
Cl^-	18	40.2	4.6	–
Na^+	10	-12.6	14.2	2.34 ± 0.06
K^+	18	4.32	16.6	2.79 ± 0.08
Rb^+	36	12.8	31.7	2.9
Cs^+	54	24.7	45.8	3.13 ± 0.07

Table 1.1: X-ray scattering contrasts of ions in water. Z is the number of electrons per ion, equal to the atomic number minus the charge, V_I is the volume per ion (i.e. the partial molar volume [38]), and Z_{eff} is calculated from Z and V_I using Equation 1.6 with $\bar{\rho}_e = 0.334\,\text{e}^-\text{Å}^{-3}$. The uncertainties are $\pm 1\,\text{Å}^3$ for V_I and $\pm 0.4\,\text{e}^-$ for Z_{eff}. R_{M-O} is the distance between the ion and the first hydration shell from Ref. [39].

1.3 Results

Duplex DNA (25 base-pair, charge of -48 e) was equilibrated with neutral pH buffers containing 100 mM monovalent salts with increasing Z; NaCl, KCl, RbCl, CsCl. X-ray scattering patterns were acquired at a relatively low DNA concentration of 50 μM duplex to minimize interparticle interference effects [40]. The intensity of each sample was normalized by the DNA concentration and placed on an absolute scale (electrons^2 per molecule) as described in Section 1.5.4. Repeated concentration measurements indicate that the normalization precision is $\pm 3\%$. Further details for sample preparation and data analysis are provided in Materials and Methods. SAXS data for the monovalent ion series (Figure 1.1a) shows an enhancement at low scattering angle for the heavy ions Rb^+ and Cs^+ relative to Na^+ and K^+, suggesting the presence of an ion atmosphere surrounding the DNA. Heavy salts increase the forward scattering contrast in a trend consistent with their increasing Z_{eff}.

The scattering data were extrapolated to $q = 0$ using an indirect Fourier transform [41]. $I(0)^{1/2}$ is plotted in Figure 1.1b as a function of the hydrated ion contrasts (Table 1.1). The linear relationship between Z_{eff} and $I(0)^{1/2}$ within experimental uncertainty is consistent with the assumption that the number of excess ions is independent of the identity of the counterion. By extrapolation of $I(0)^{1/2}$ to the point where the net ion contrast is zero (Figure 1.1b), we

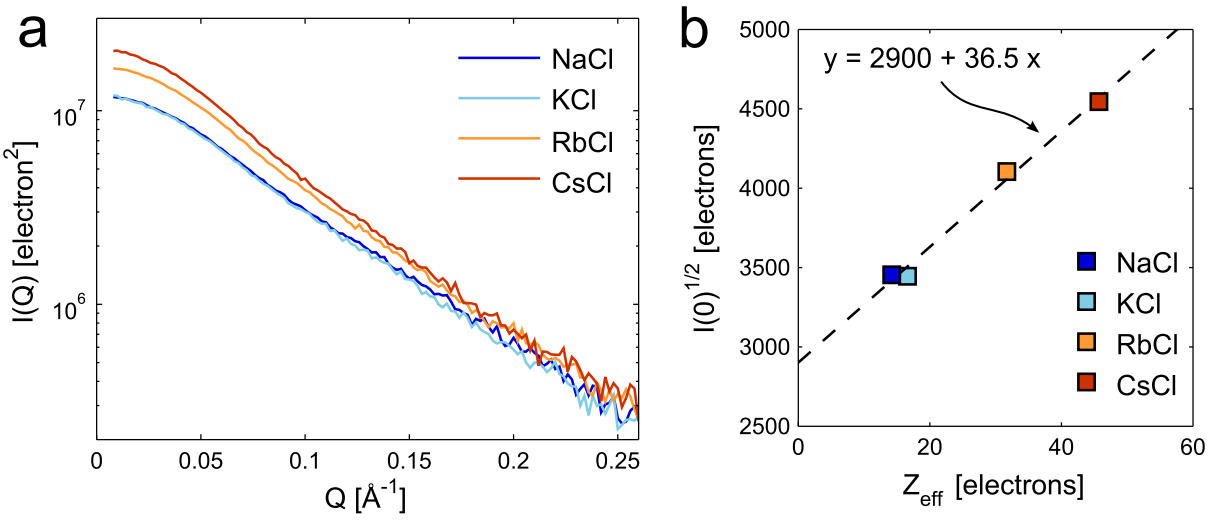

Figure 1.1: Small angle x-ray scattering from 25 base-pair DNA with monovalent salts. (a) The background-subtracted scattering intensity $I(q)$ vs. momentum transfer q is shown an absolute scale (electron^2 per DNA molecule). The scattering intensity at low angles depends on the identity of the counterion, indicated in the legend. (b) The extrapolated forward scattering amplitude, $I(0)^{1/2}$, varies linearly with the effective ion contrast, $Z_{\rm eff} = Z - \bar{\rho}_s V_{\rm ion}$, where $\bar{\rho}_s$ is the solvent electron density and V_{ion} is the change in solution volume due to the ion. When interpreted as a heavy atom substitution experiment (see text), the slope is equal to the number of excess ions: $N_{\rm I} = 36.5 \pm 2.4$, and the intercept is the contrast contribution of the DNA, its hydration water, and excluded volume; $\delta_M = 2900 \pm 50$ electrons.

obtain the DNA contrast according to Equation 1.7, $\delta_M = 2900 \pm 50$. From the slope, a number of excess ions $N_I = 36.5 \pm 2.4$ is obtained. The number of Rb^+ ions was also measured using the two-energy ASAXS methods described in Ref. [31] (data shown in Figure 1.4a). We find $N_I = 38.9 \pm 2.5$, in good agreement with the heavy atom technique. Since the DNA charge is $-48e$ at neutral pH, the measured values of N_I indicate that $75-80\%$ of the charge is neutralized by monovalent counterions, in accord with previous equilibrium dialysis measurements that found $80 \pm 4\%$ neutralization by excess Na^+ around a DNA duplex of similar length and composition in 100 mM NaCl [18]. The remaining DNA charge is neutralized by co-ion exclusion [18, 42]. The low x-ray contrast of hydrated Cl^- (Table 1.1) means that the heavy-atom measurement is relatively insensitive to co-ions. The exclusion of ~11 Cl ions has the predicted result of depressing $I(0)^{1/2}$ by approximately 50 electrons, which is within the measurement uncertainty.

Next, the heavy atom SAXS data were decomposed into components using Equation 1.12. To provide quantitative comparison between experimentally-determined terms $P_{exp}(q)$ and those predicted by models $P_{calc}(q)$, we compute the following chi-squared statistic for each pair of curves:

$$\chi^2_{rel} = \frac{1}{\beta^2 N} \sum_{i=1}^{N} \left(\frac{P_{exp}(q_i) - P_{calc}(q_i)}{\sigma_{exp}(q_i)} \right)^2$$

where σ_{exp} is an estimate of the experimental uncertainty and β is a scale factor so that $\chi^2_{rel} = 1$ when $P_{calc}(q)$ is replaced by the regularized version of $P_{exp}(q)$ (smoothed data in Figures 1.2 and 1.3). Although in the ideal case $\beta = 1$, we find $\beta \sim 0.94$ when σ_{exp} is determined by propagation of uncertainty through Equation 1.12.

In Figure 1.2a, the experimentally-determined DNA component is compared with a theoretical prediction for hydrated DNA where the measured contrast δ_M constrained DNA's hydration parameters N_H and N_E according to Equation 1.8, enabling a predictive calculation of DNA's scattering profile (note that programs such as CRYSOL fit the hydration parameters to the shape of the experimental scattering curve on an arbitrary intensity scale [43]). The procedure for computing the scattering amplitude of the hydrated DNA is illustrated in Figure 1.2b and described in Section 1.5.6. The calculated profile and the raw data agree within experimental uncertainty ($\chi^2_{rel} = 1.14$).

The experimentally-determined cross-term $P_{MI}(q)$ and ion-only term $P_I(q)$ are shown in Figure 1.3a and b, respectively. Predicting these two terms using the NLPB equation requires a choice of Stern layer thickness r, nominally equal to the radius of the hydrated counterion. Radii for hydrated ions considered here range from $3.8 - 4.5\,\text{Å}$ (R_{M-O} plus $1.4\,\text{Å}$ for the water molecule radius: see Table 1.1 and Ref. [18]). However, previous studies using NLPB have assumed $r = 2\,\text{Å}$ [20]. Therefore, we performed calculations using both $r = 2\,\text{Å}$ and $r = 4\,\text{Å}$. We find that the shape of the predicted cross-term shown in Figure 1.3a depends on the value of r. Comparing data and calculation, $r = 2\,\text{Å}$ shows a slight improvement over $r = 4\,\text{Å}$ ($\chi^2_{rel} = 1.28$

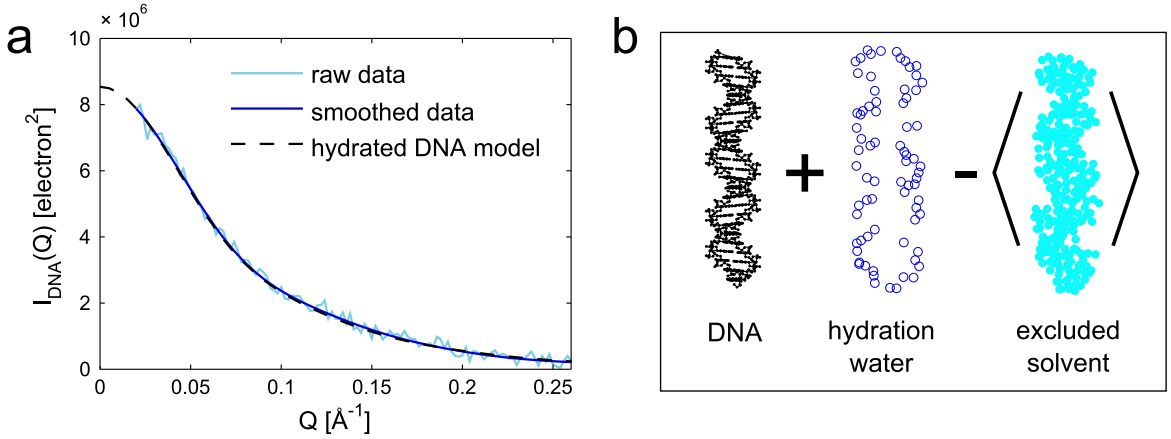

Figure 1.2: (a) The ion-independent (DNA-only) component of the scattering was found from decomposition of the scattering profiles in Figure 1.1a (raw data), and from decomposition of the same profiles with prior smoothing by regularization [41] (smoothed data). The theoretical scattering intensity for hydrated DNA shown in (a) is the absolute square of the amplitude, whose components are illustrated in (b). Details of this calculation are provided in Section 1.5.6. Briefly, a B-form DNA model was generated by Nucleic Acid Builder [44], and water molecules were placed uniformly at the surface to approximate the hydration layer. The excluded solvent was modeled using a pre-computed water box. The numbers of hydrating and excluded waters were constrained by the measured DNA contrast δ_M (see Equation 1.8). For clarity, only water molecules within a 5 Å slab centered on the DNA are shown in (b).

for $r = 2$ Å vs. $\chi^2_{rel} = 1.92$ for $r = 4$ Å). For both values of r, $P_{MI}(q)$ computed by NLPB decays more rapidly than the experimental curve. The width of the cross-term is inversely related to the distance between DNA and its counterions; thus the NLPB ion distribution appears more extended than the experiments suggest. Stern layer-dependent differences are also observed in the calculated ion-only term $P_I(q)$, but the experimental $P_I(q)$ is too noisy to discriminate between models ($\chi^2_{rel} = 1.01$ for $r = 2$ Å and $\chi^2_{rel} = 1.02$ for $r = 4$ Å).

The three-term decomposition of the heavy atom data assumes that the ion distributions are the same in the different salts used. However, ASAXS is measured using a single sample. The anomalous difference for two energies is a mixture of the cross-term and the ion-only terms [25]:

$$I_{lo}(q) - I_{on}(q) \quad = \quad 2\delta_M N_I \left(f'_{lo} - f'_{on}\right) \left(P_{MI}(q) + \gamma P_I(q)\right) \tag{1.9}$$

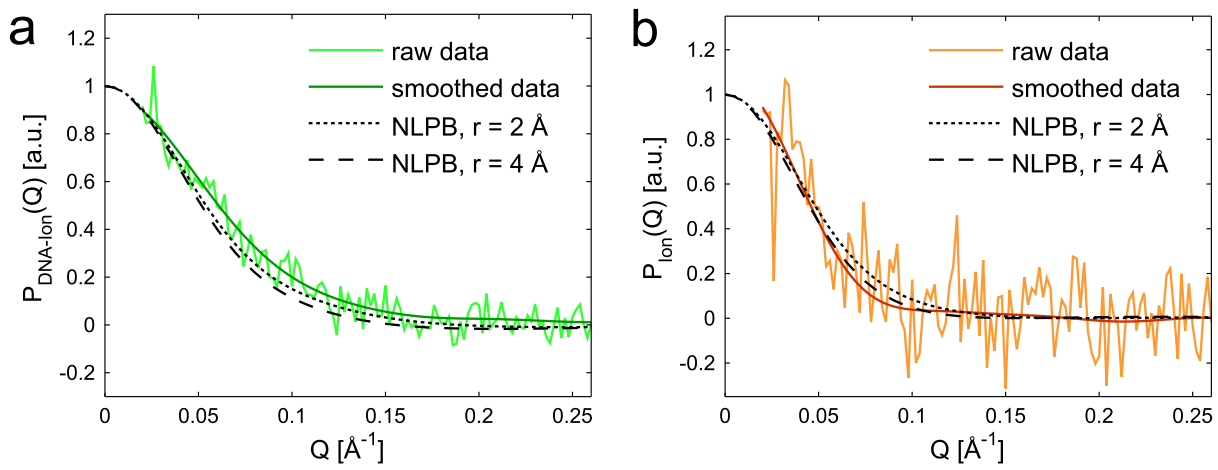

Figure 1.3: Partial scattering intensities for the DNA-ion cross term (a) and the ion-only term (b) were obtained from a decomposition of the data in Figure 1.1a (raw data), and with prior smoothing by regularization [41] (smoothed data). Also shown are Nonlinear Poisson-Boltzmann equation (NLPB) predictions with two values of the Stern layer thicknesses, r, indicated in the legend.

where $\gamma = N_I\left(Z_{\text{eff}} + 0.5\left(f'_{lo} + f'_{on}\right)\right)\delta_M^{-1}$ and f'_{lo} and f'_{on} are the real parts of the anomalous scattering factor for Rb^+ at each energy (for numerical values see Figure 1.4a). The NLPB predictions for $P_{MI}(q)$ and $P_I(q)$ are compared with the experimental anomalous difference signal for DNA in 100 mM RbCl, shown in Figure 1.4. Again, the agreement with NLPB is poor for $r = 4$ Å ($\chi^2_{rel} = 15.4$), and improved for $r = 2$ Å ($\chi^2_{rel} = 1.65$). To provide a cross-check for the basis functions determined by the heavy atom method, we also computed the anomalous difference using the experimental functions $P_{MI}(q)$ and $P_I(q)$ in Figure 1.3. This anomalous difference profile is plotted in Figure 1.4. Agreement between the two experimental methods is excellent ($\chi^2_{rel} = 1.39$).

Thus, ASAXS and heavy-atom experiments provide a self-consistent measurement of the monovalent ion atmosphere around DNA. Both techniques suggest that the ions are closer to the DNA than predicted by NLPB. In order to gain additional insight into the origin of this discrepancy, we used NLPB and the hydrated DNA model to predict the full SAXS profiles. The experimental SAXS curves have a much higher signal-to-noise ratio than the decomposed

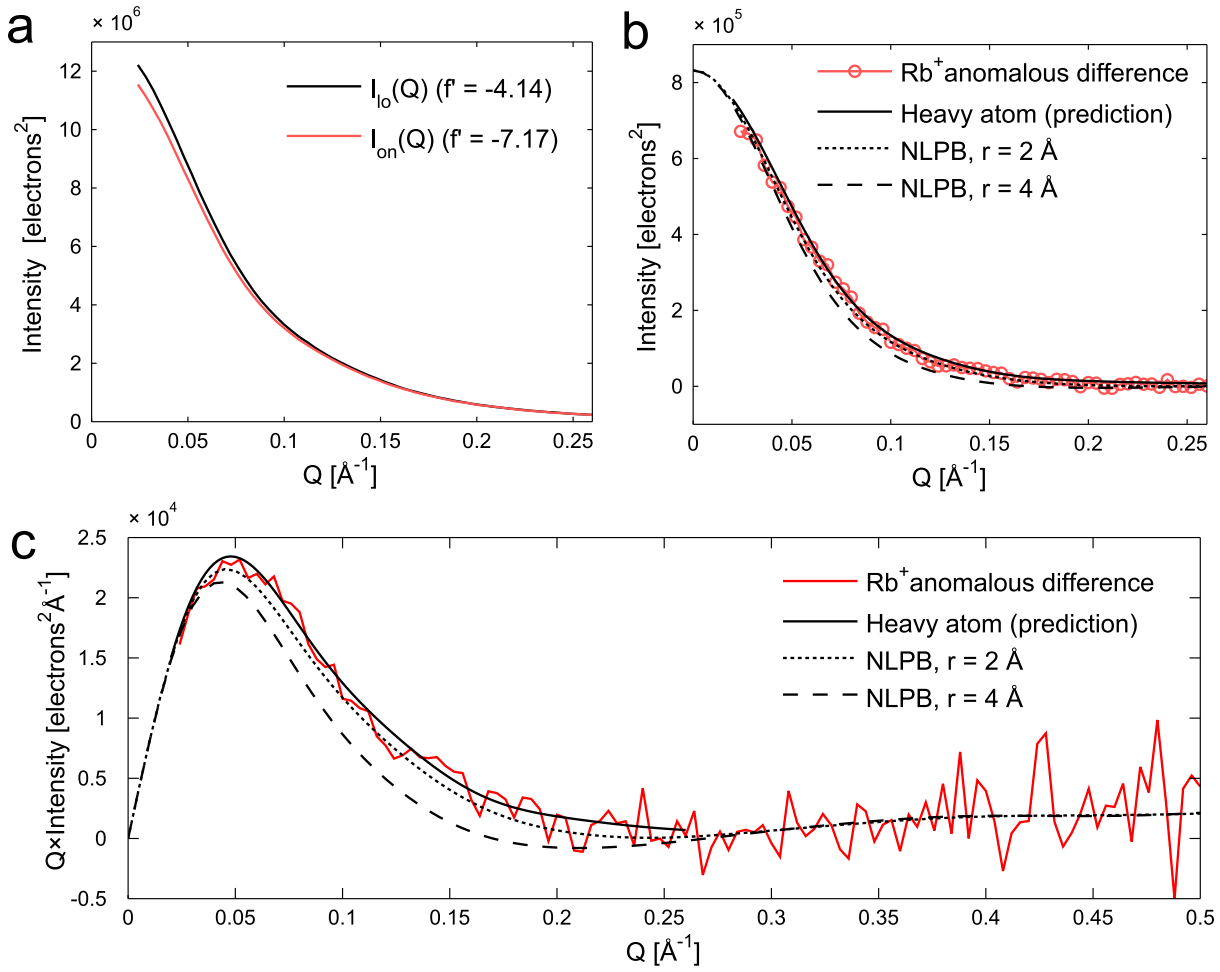

Figure 1.4: Anomalous x-ray scattering (ASAXS) data for DNA with Rb^+ counterions. (a) Scattering profiles were measured on an absolute scale at two energies below the absorption edge of Rb. The real part of the anomalous scattering factor of Rb^+ is given in the legend at each energy (labeled *lo* and *on*). (b) The anomalous difference between curves in (a), $I_{lo}(q) - I_{on}(q)$, is compared with predictions from NLPB theory and with a curve synthesized from the experimental basis functions in Figure 1.3 (see Equation 1.9). (c) The profiles in (b) are plotted on $q \times I(q)$ vs. q axes to facilitate visual comparison.

scattering curves, and are consequently a more challenging test of NLPB. SAXS profiles corresponding to the hydrated DNA model and an ion atmosphere from NLPB with $r = 2\,\text{Å}$ are shown in Figure 1.6a (dark blue curves). Agreement between experimental data and NLPB predictions is poor, especially for the heavy counterions Cs^+ and Rb^+ where the ion atmosphere represents a significant proportion of the total scattering.

We hypothesize that NLPB underestimates the number of counterions present at the DNA surface. This counterion deficit was suggested by simulations of ions around RNA that showed a stronger ion-surface interaction than predicted by NLPB [21]. Computational approaches more sophisticated than NLPB, such as molecular dynamics (MD) [45] and integral equation theory [36] could be similarly compared with x-ray data using the methodology presented here. We note that favorable comparisons between explicit solvent MD and ASAXS data for Rb^+ ions around an RNA duplex were obtained previously [26]. However, for the purposes of this study, we leave the question of the energetics of ion-DNA association to future work, and focus on the spatial distribution of ions.

To test the surface-association hypothesis, we model the ion distribution in an approximate but realistic way commensurate with the low resolution of the SAXS measurement. Although MD has shown significant disagreement with NLPB near the molecular surface, NLPB does an adequate job describing the diffuse ion atmosphere [45]. Therefore, as a first-order correction, we placed explicit counterions on the surface of DNA, and used NLPB to model the remaining ions. The procedure for placing counterions was as follows. First, an ion was placed at the site with the greatest magnitude of the electrostatic potential according to the NLPB solution. The NLPB equation was re-solved with this ion present, the new potential guided the placement of a second ion. This procedure was repeated, building up progressively more ions around the DNA: models with 0, 5, 10, 15, and 20 ions are shown in Figure 1.5a. For each of these models, the ion distribution including explicit and NLPB-modeled ions is shown in Figure 1.5b. As

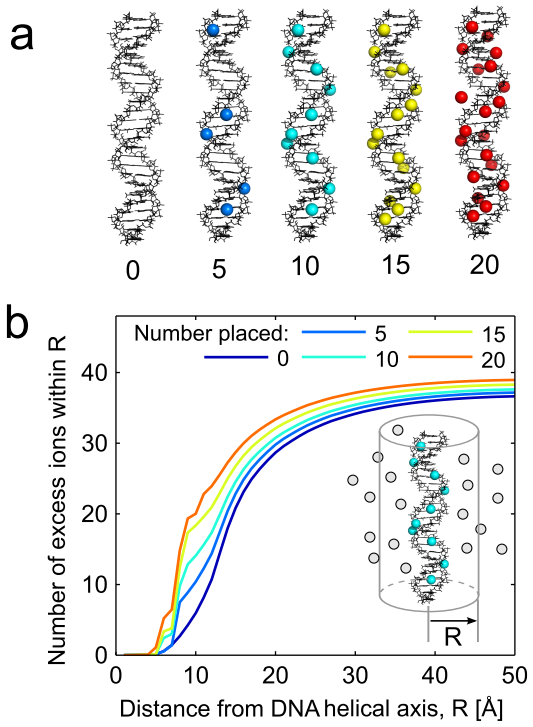

Figure 1.5: Modeling of surface-localized ions and their effect on the diffuse ion atmosphere. (a) Ions were placed sequentially at sites on the surface of the DNA molecule with the greatest magnitude of the potential, and the NLPB equation was re-solved after each ion addition (see text). DNA models with 0, 5, 10, 15, and 20 ions (colored spheres with $r = 2\,\text{Å}$) were rendered using Pymol version 1.2r1 (DeLano Scientific LLC). The ions are found primarily in the minor grooves. (b) The number of excess ions within a cylinder of radius R centered on the DNA (inset) was computed for each model. The number of excess ions includes the explicit ions shown in (a) and the excess ion density from the NLPB solution (represented as gray circles in the inset).

ions are added to the surface, the number of diffuse ions predicted by NLPB decreases, making the total number of excess ions only weakly dependent on the number explicitly added. The total number of excess ions within $R = 50\,\text{Å}$ from the central axis of the DNA helix ranges from 36.7 (no surface ions) 39.0 (20 surface ions). These numbers are currently within the uncertainty of ion-counting methods. However, the shape of the ion atmosphere depends on the number of ions at the surface, and SAXS may discriminate between these models according to the predicted scattering profiles.

In Figure 1.6a, experimental SAXS data are compared with predicted profiles for each ion

atmosphere model in Figure 1.5. From visual comparison of theory and experiment for DNA in CsCl and RbCl salts, the agreement improves and then worsens with increasing numbers of surface-localized ions, with a best-fit value between 10 and 15. This is also shown quantitatively in Figure 1.6b using a chi-squared statistic. The residuals for the data and prediction with 10 surface ions in Figure 1.6c are essentially flat for $q \gtrsim 0.04\,\text{Å}^{-1}$. The non-zero residuals at the lowest scattering angles, corresponding to length-scales comparable to the DNA length, may be a consequence of using an ideal B-form model, of assuming a uniform hydration layer, or of a small amount of interparticle interference in the data despite the low concentration of 50 μM. However, varying the number of surface-localized ions in the model affects scattering data in the mid-q region (length scales comparable to the DNA diameter) so there is little danger of over-fitting the low angle data. Intriguingly, the chi-squared minima in Figure 1.6b depend on cation identity, with an order $Cs^+ < Rb^+ < K^+ < Na^+$ that correlates inversely with the radius R_{M-O} and partial molar volume V_I of the cations (see Table 1.1). However, considering the uncertainty introduced by the modeling procedure and the experimental noise, the present data are consistent with an identical ion distribution for all cations (i.e. distributions with $10-15$ surface ions).

1.4 Discussion

We develop methods for collecting and analyzing SAXS data using the contrast variation technique of heavy atom replacement. These methods build on previous experience with ASAXS by adding absolute intensity calibration to the heavy atom measurement. We find that the scattering contrasts of monovalent ions around DNA are well-described by predictions based on their partial molar volumes, and obtain excess ion numbers consistent with previous measurements. In addition, the number of water molecules displaced by DNA is measured by extrapolation to zero ion contrast. Heavy atom data are decomposed into contributions from the ions and DNA

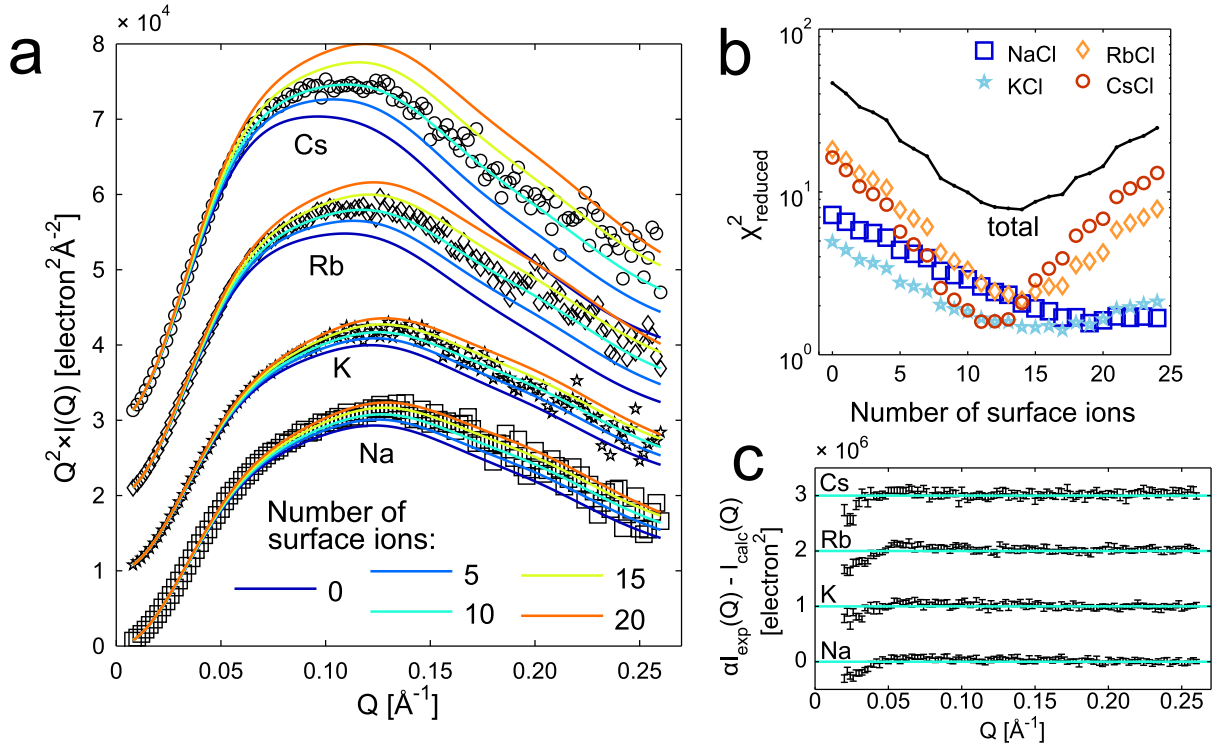

Figure 1.6: Adding surface-localized ions to DNA models improves the fit with experimental data. (a) Theoretical scattering profiles were computed from the hydrated DNA model (Figure 1.3a,d) and an ion atmosphere modeled using the NLPB equation (dark blue), and using the NLPB equation with explicit ions at the DNA surface, shown in Figure 1.5a. Profiles were computed for each ion type (Na, K, Rb, and Cs), plotted as $I(q) \times q^2$ to emphasize the high-angle features, and offset by 1×10^4 according to ion type for clarity. These are compared with experimental intensities (Figure 1.1a) using a scale factor α to minimize the χ^2 for each predicted curve. To aid in visual comparison, the experimental data in (a) were scaled to match the theoretical profiles computed using 10 surface ions. The scale factor α deviated from 1 by less than 3%. (b) The agreement between experimental data and each theoretical profile was assessed using χ^2. The reduced χ^2 value is shown on a logarithmic scale for each experimental curve (NaCl, KCl, RbCl, CsCl) and for the whole data set (total). (c) The residual with experimental error bars for 10 surface ions is plotted in on a standard intensity scale in, with each curve offset by 1×10^6 for clarity.

alone. The DNA-only profile agrees with predictions from a duplex DNA model with hydration parameters set according to the experimentally measured number of water molecules displaced. However, when the NLPB equation is used to model the counterion atmosphere, we find significant disagreement between predicted x-ray scattering profiles and three related measurements: (1) the ion-DNA cross-term determined by heavy-atom substitution; (2) the ASAXS anomalous difference profiles; and (3) SAXS profiles for DNA with electron-dense counterions. All of these measurements are sensitive to the spatial distribution of ions around DNA and suggest that the counterions are more closely localized to the DNA surface than predicted.

Our conclusion that NLPB does not correctly model the monovalent ion atmosphere around DNA is seemingly at odds with a number of previous x-ray scattering studies that demonstrate agreement between NLPB and experiment [19, 23, 27, 46]. However, there are important differences between the present test of NLPB and previous attempts. First, we calculate the ion distribution using an all-atom representation of the DNA molecule, a clear improvement over previous studies that modeled DNA as a charged cylinder [23, 27]. In a previous ASAXS paper, we also solved NLPB using an all-atom DNA representation but reported only a qualitative comparison between theory and experiment [46]. In a single prior study, we compared X-ray scattering data directly with all-atom NLPB calculations [19]. In that study, SAXS and ASAXS data for a short DNA duplex in a variety of salts (including counterions studied here) compared favorably with predicted profiles [19]. The most important difference between this previous study and the present one is the ionic strength I of the solution: here, we used $I = 100$ mM, whereas $I = 400$ mM was used in the previous study. The ionic strength affects the length scale of the diffuse ion atmosphere through the Debye screening length, which is proportional to $I^{-1/2}$. We speculate that previous SAXS measurements were unable to distinguish between surface-localized and diffuse counterions because the diffuse ion atmosphere was too compact (screening length $\sim 5\,$Å) relative to the resolution of SAXS. In the low ionic strength measured here (screening length $\sim 10\,$Å), the spatial separation between diffuse and surface-localized

ions is more pronounced, and therefore our measurement is more sensitive to the relative populations.

Although x-ray scattering data for DNA have not previously shown disagreement with NLPB, other techniques have revealed shortcomings of the theory. For example, counterion competition experiments using equilibrium dialysis show that NLPB does not account for the competition of Mg^{2+} against a background of 20 mM Na^{+} [18], even when NLPB is modified to account for the unequal ion sizes [20]. Furthermore, NLPB does not reproduce the monovalent ion distributions seen in explicit solvent MD simulations of RNA duplexes [21]. Interestingly, the counterion competition experiments and MD results are both consistent with the hypothesis that NLPB underestimates the free energy for ions near the surface of nucleic acids. However, the details of this ion-DNA interaction are not well understood.

There are several simplifying assumptions in NLPB theory that may explain its inaccuracy. For example, NLPB neglects the possibility of direct interactions between ions and DNA involving full or partial dehydration of the ion [2]. Such dehydration interactions are unlikely to be important for the monovalent ions studied here, as NMR measurements suggest that Group I cations interact with DNA in a fully-hydrated state [22], and the physiologically important divalent ion Mg^{2+} is usually fully-hydrated when observed in RNA crystal structures (with rare exceptions that have functional relevance) [2]. To explain the discrepancy between NLPB and counterion-competition data, Chu et al. raise the possibility that ion-ion correlations, neglected in the mean-field NLPB theory, may be important [20]. Finally, NLPB neglects solvent degrees of freedom, treating water as a uniform dielectric medium. Intriguingly, Kirmizialtin and coworkers recently showed that artificially lowering the solvent dielectric constant near the surface in NLPB calculations dramatically improves agreement with explicit solvent MD simulations [21].

We conclude that under physiological conditions, the ion atmospheres around DNA are

strongly influenced by the electrostatic screening properties of the solvent near the highly charged molecular surface. New integral equation theories that account for solvent degrees of freedom promise to provide the realism of molecular dynamics without the computational cost [36]. The x-ray scattering methods presented here provide a unique measurement of the global structure of the ion atmosphere, and will be essential for testing computational methods for efficient and accurate modeling of electrostatic interactions between charged biological molecules.

Acknowledgments

We thank CHESS scientists Arthur Woll and Ken Finkelstein and CHESS staff for assistance with the experiments, members of the Pollack Lab for help with x-ray data collection, and David Case, Hung Nguyen, Nathan Baker, Serdal Kirmizialtin, and Ron Elber for insightful discussions. This work was supported by the NIH (R01-GM085062). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0936384, and the MacCHESS resource is supported by the National Institute of General Medical Services (NIGMS) award GM-103485.

1.5 Materials and Methods

1.5.1 Sample preparation

The 25 base-pair DNA duplex was composed of two single stranded oligonucleotides with sequence GCATCTGGGCTATAAAAGGGCGTCG (S1) and its complement (S2). The sequence has been used in many previous SAXS studies [3, 4, 19, 27, 31, 32, 46–48]. Oligonucleotides were synthesized and HPLC purified by Integrated DNA Technologies (Coralville, IA) and de-

livered as lyophilized powders. Each strand was rehydrated in aqueous buffer containing 10 mM TRIS, 50 mM NaCl, 1 mM EDTA, pH 8.0. The concentration of each strand was calculated from the UV absorption at 260 nm using extinction coefficients derived from the nearest-neighbor model [49]; $\varepsilon_{S1} = 244{,}400\,\mathrm{L\,mol^{-1}cm^{-1}}$, and $\varepsilon_{S2} = 228{,}500\,\mathrm{L\,mol^{-1}cm^{-1}}$. Strands were mixed in an equimolar ratio, annealed at 94 C for 4 minutes, and allowed to cool on the bench.

Buffered salt solutions were prepared using 1 mM Na-MOPS pH 7.0, 100 mM of 1:1 salt (NaCl, KCl, RbCl, CsCl) and NANOpure water (Barnstead, Dubuque, IA). All reagents were purchased from Sigma-Aldrich, unless specified. DNA in each buffered salt solution was prepared by spin dialysis using Amicon Ultra-0.5 mL, 10 kDa cutoff columns (Millipore, Billerica, MA), repeated to ensure complete exchange, and the final flow-through was retained for SAXS background subtraction. The UV absorbance of each duplex DNA solution at 260 nm was converted to concentration using the extinction coefficient of each strand with hypochromicity corrections [50], $\varepsilon_{S1+S2} = 397{,}558\,\mathrm{L\,mol^{-1}cm^{-1}}$. Samples were diluted with the matching buffer to a concentration of 50 μM (SAXS) and 150 μM (ASAXS).

1.5.2 X-ray data collection

SAXS data were collected at the Cornell High Energy Synchrotron Source (CHESS) beamline G1, which has a 49-pole wiggler source and multilayer optics. The monochromator was tuned to an x-ray energy of 10.53 keV, and the beam size was defined using slits to $300\,\mu\mathrm{m\,(H)} \times 250\,\mu\mathrm{m\,(V)}$. Guard slits upstream of the sample blocked parasitic scattering from the beam-defining slits, and the x-ray path between the beam-defining slits and the detector was evacuated (~10 mTorr) to reduce the background. The variability in the x-ray intensity between exposures was normalized using the photocurrent from a PIN diode integrated into the beamstop. Samples

were held in an in-vacuum quartz capillary with a 2 mm diameter and 10 μm wall thickness (Hampton Research, Aliso Viejo, CA). The liquid sample was oscillated through the beam using a computer-controlled syringe pump to mitigate radiation damage [51]. Scattering patterns were recorded with a photon-counting area detector (Pilatus 100K, Dectris) at distance of 1.689 m from the sample position (q-range of 0.008 to 0.260 Å^{-1}) in four 60s exposures.

The ASAXS experiments were performed at CHESS' C-line using bend-magnet radiation and a double-bounce Si-110 monochromator. The sample was held in a 3 mm diameter quartz capillary and oscillated during x-ray exposure. An evacuated flight tube with a Si_3N_4 window on the upstream end was positioned between the sample capillary and detector, both of which were in air. Images were recorded on the area detector (Pilatus 100K) at a distance of 0.957 m from the sample (q-range of 0.024 to 0.500 Å^{-1}). A semi-transparent beamstop consisting of a stack of Mo foils (Goodfellow) with a total thickness of $\sim 550\,\mu$m was mounted inside the flight-tube. The small fraction of the primary beam that passed through the beamstop was recorded on the detector, and the centroid was defined as $q = 0$ for azimuthal integration. A Si drift detector (X-flash, Roentec) recorded the x-rays elastically scattered by the beamstop for normalization of the SAXS profiles. Scattering data were acquired at two energies below the Rb^+ K-edge, 15.093 keV (lo) and 15.191 keV (on), cycling between energies every 2 minutes, with a total exposure time of 32 minutes per sample.

1.5.3 X-ray data processing

SAXS and ASAXS data were processed in Matlab (MathWorks, Natick, MA) using routines developed in-house. The detector coordinates were converted to momentum transfer ($q = 4\pi \sin(2\theta/2)/\lambda$, where 2θ is the scattering angle, and λ is the wavelength) using a silver behenate powder diffraction standard [52]. Pixels occluded by the beamstop were masked,

and the 1-D intensity was computed from each 2-D image by binning pixels according to the magnitude of q, with constant bin widths of $0.002\,\text{Å}^{-1}$ (SAXS) and $0.004\,\text{Å}^{-1}$ (ASAXS). The uncertainty of $I(q_i)$ was estimated using $\Delta I(q_i) \approx \sigma_i N_i^{-1/2}$, where σ_i is the standard deviation among the N_i pixels in the bin. This uncertainty was propagated through subsequent averaging and buffer subtraction operations. To verify that no significant radiation damage occurred during the experiment, consecutive exposures of each sample were inspected for time-dependence before averaging them together.

1.5.4 Absolute intensity calibration

For the ASAXS and heavy atom experiments, the scattering intensity was placed on an absolute scale using liquid water as a calibrant [53]. The forward x-ray scattering of a liquid (i.e. the macroscopic scattering cross-section, $d\Sigma/d\Omega$) is proportional to the isothermal compressibility:

$$\frac{d\Sigma}{d\Omega} = n\, r_0^2\, Z^2\, k_B T\, \chi_T$$

where r_0 is the classical electron radius, n is the molecular number density, Z is the number of electrons per molecule (10 for water), and χ_T is the (temperature-dependent) isothermal compressibility, which has been determined accurately using speed of sound measurements [54]. For liquid water at 23C, $\chi_T = 4.55 \times 10^{-10}\text{Pa}^{-1}$ and $d\Sigma/d\Omega = 0.0164\,\text{cm}^{-1}$.

For calibrating x-ray data, the scattering curve for pure water was measured in the sample cell, and the scattering from the empty cell was measured and subtracted. The scattering curves for each macromolecular sample, $I(q)_{\text{sample}}$, were scaled by a constant:

$$\left.\frac{d\Sigma(q)}{d\Omega}\right|_{\text{sample}} = \frac{I(q)_{\text{sample}}}{I(q \to 0)_{\text{water}}} \left.\frac{d\Sigma}{d\Omega}\right|_{\text{water}}$$

The absolute scattering cross section was converted to units of electron^2 per molecule by dividing by the concentration of molecules (number density n) and the square of the classical

electron radius r_0:

$$I(q)\ \left[\text{electron}^2\right] = \frac{1}{n r_0^2}\frac{d\Sigma(q)}{d\Omega}$$

1.5.5 Method for decomposing experimental heavy atom data

Equation1.3 written in matrix form, $A = B \cdot C$, becomes:

$$\begin{bmatrix} I^{(1)}(q_1) & I^{(2)}(q_1) & \cdots \\ I^{(1)}(q_2) & I^{(2)}(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} P_M(q_1) & P_{MI}(q_1) & P_I(q_1) \\ P_M(q_2) & P_{MI}(q_2) & P_I(q_2) \\ \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} \delta_M^2 & \delta_M^2 & \cdots \\ 2\delta_M N_I \delta_I^{(1)} & 2\delta_M N_I \delta_I^{(2)} & \cdots \\ N_I^2\left(\delta_I^{(1)}\right)^2 & N_I^2\left(\delta_I^{(2)}\right)^2 & \cdots \end{bmatrix} \tag{1.10}$$

where superscripts enumerate the different salt solutions. In the case where the matrix A contains experimental data, Equation 1.10 is

$$A_{exp} = BC + E$$

where E is a matrix of errors due to measurement noise. The solution B that minimizes the least-squares error is

$$B = A_{exp} C^{+}$$

where C^{+} is the Moore-Penrose inverse of C. However, finding B in this way will only work with perfectly-calibrated data. The solution can be found more robustly if both sides of Equation 1.10 are normalized by the forward scattering:

$$A_{exp} N_{exp}^{-1} = B\left(CN^{-1}\right) \tag{1.11}$$

where N_{exp} is a diagonal matrix containing the experimentally-determined zero-angle scattering

$$N_{exp} = \begin{bmatrix} I^{(1)}(0) & 0 & \cdots \\ 0 & I^{(2)}(0) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

and N is the prediction for N_{exp},

$$N = \begin{bmatrix} \left(\delta_M + N_I \delta_I^{(1)}\right)^2 & 0 & \cdots \\ 0 & \left(\delta_M + N_I \delta_I^{(2)}\right)^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

The robust least-squares solution for B in Equation 1.11 is

$$B = A_{exp} N^{-1} \left(CM^{-1}\right)^+ \tag{1.12}$$

where $\left(CM^{-1}\right)^+$ is the Moore Penrose inverse of (CM^{-1}).

1.5.6 Calculation of SAXS profiles

We modeled the DNA as an ideal B-form duplex prepared using Nucleic Acid Builder [44]. Hydration effects were added using a geometric method similar to the program HYDCRYST [55]. First, water molecules were placed uniformly around the DNA at a distance of $1.50 - 1.82$Å from the surface ($N_H = 348$, a ratio of 8 waters per phosphate group), where the molecular surface is defined by the effective Van der Waals radii of the atomic groups [43]. Scattering amplitudes were computed following a method used for molecular dynamics simulations [56], illustrated in Figure 1.2b. The scattering amplitude of the hydrated DNA is

$$A_M(q) = A_{DNA}(q) + A_H(q) - \langle A_E(q) \rangle_\Pi \tag{1.13}$$

where the amplitudes $A_{DNA}(q)$, $A_H(q)$, and $A_E(q)$ correspond to the DNA (in-vacuum), the hydration waters, and the excluded solvent, respectively. Brackets denote the configurational average. The excluded solvent was modeled using a pre-computed water box [57], spatially filtered according to the DNA model. Because solvent is excluded by both DNA and its hydration layer, the molecular surface was expanded until the total number of excluded solvent particles equaled the value of N_E according to Equation 1.8 with the measured DNA contrast δ_M. The

average scattering amplitude from excluded solvent was approximated using a spherical Gaussian form-factor for each solvent particle [43, 56], and the configuration average was computed for eight positions of the DNA within the water box.

To predict the counterion densities around DNA, we solved the Poisson-Boltzmann equation numerically on a rectangular grid using the program APBS [17]. The DNA was placed in the center of a $100\,\text{Å} \times 100\,\text{Å} \times 168\,\text{Å}$ box with a uniform mesh of $129 \times 129 \times 193$ elements. A zero-potential boundary condition was used. The excess ion density was sampled at 10,000 discrete points, where the probability of placing a point randomly with a grid cell was proportional to the excess number density at that position. The amplitudes were computed efficiently using the spherical harmonic expansion of the coordinates [43]. The q-dependence of the ion scattering factors were neglected: an accurate approximation because $P_I(q)$ and $P_{MI}(q)$ decay rapidly. The fact that the real ion atmosphere is a discrete rather than continuous density does not affect the DNA-ion cross term because discrete effects average out (see Equation 1.1). However, discrete effects are potentially important in the ion-only term because the amplitude is squared prior to averaging. As a first-order correction, one could sample the NLPB ion density at only N_I points and average the intensity for multiple ion configurations. An equivalent and more computationally efficient correction is: $P_I(q) \leftarrow P_I(q) \times (1 - N_I^{-1}) + N_I^{-1}$. We applied this correction to the NLPB ion-only term for completeness; it does not affect the conclusions of this work.

References

[1] C. Garcıa-Garcıa and D. E. Draper, "Electrostatic interactions in a peptide–RNA complex", J. Mol. Biol. **331**, 75–88 (2003).

[2] D. E. Draper, "A guide to ions and RNA structure", RNA **10**, 335–343 (2004).

[3] X. Qiu, L. W. Kwok, H. Y. Park, J. S. Lamb, K. Andresen, and L. Pollack, "Measuring Inter-DNA potentials in solution", Phys. Rev. Lett. **96**, 138101+ (2006).

[4] X. Qiu, K. Andresen, J. S. Lamb, L. W. Kwok, and L. Pollack, "Abrupt transition from a free, repulsive to a condensed, attractive DNA phase, induced by multivalent polyamine cations", Phys. Rev. Lett. **101**, 228101 (2008).

[5] X. Qiu, V. A. Parsegian, and D. C. Rau, "Divalent counterion-induced condensation of triple-strand DNA", Proc. Natl. Acad. Sci. U.S.A. **107**, 21482–21486 (2010).

[6] V. A. Bloomfield, "DNA condensation by multivalent cations", Biopolymers **44**, 269–282 (1997).

[7] S. J. Tan, J. S. Kahn, T. L. Derrien, M. J. Campolongo, M. Zhao, D.-M. Smilgies, and D. Luo, "Crystallization of DNA-capped gold nanoparticles in high-concentration, divalent salt environments", Angew. Chem. **126**, 1340–1343 (2014).

[8] I. Y. Wong and N. A. Melosh, "Directed hybridization and melting of DNA linkers using counterion-screened electric fields", Nano Lett. **9**, 3521–3526 (2009).

[9] M. H. Abouzar, A. Poghossian, A. G. Cherstvy, A. M. Pedraza, S. Ingebrandt, and M. J. Schöning, "Label-free electrical detection of DNA by means of field-effect nanoplate capacitors: experiments and modeling", Physica Status Solidi A **209**, 925–934 (2012).

[10] A. Poghossian, A. Cherstvy, S. Ingebrandt, A. Offenhäusser, and M. Schöning, "Possibilities and limitations of label-free detection of DNA hybridization with field-effect-based devices", Sensor. Actuat. B-Chem. **111**, 470–480 (2005).

[11] E. Stern, R. Wagner, F. J. Sigworth, R. Breaker, T. M. Fahmy, and M. A. Reed, "Importance of the Debye screening length on nanowire field effect transistor sensors", Nano Lett. **7**, 3405–3409 (2007).

[12] A. Star, E. Tu, J. Niemann, J.-C. P. Gabriel, C. S. Joiner, and C. Valcke, "Label-free detection of DNA hybridization using carbon nanotube network field-effect transistors", Proc. Natl. Acad. Sci. U.S.A. **103**, 921–926 (2006).

[13] G. S. Manning, "Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties", J. Chem. Phys. **51**, 924–933 (1969).

[14] P. Grochowski and J. Trylska, "Continuum molecular electrostatics, salt effects, and counterion binding–a review of the Poisson-Boltzmann theory and its modifications", Biopolymers **89**, 93–113 (2008).

[15] W. Rocchia, E. Alexov, and B. Honig, "Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions", J. Phys. Chem. B **105**, 6507–6514 (2001).

[16] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: application to microtubules and the ribosome", Proc. Natl. Acad. Sci. U.S.A. **98**, 10037–10041 (2001).

[17] M. Holst, N. Baker, and F. Wang, "Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples", J. Comput. Chem. **21**, 1319–1342 (2000).

[18] Y. Bai, M. Greenfeld, K. J. Travers, V. B. Chu, J. Lipfert, S. Doniach, and D. Herschlag, "Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids", J. Am. Chem. Soc. **129**, 14981–14988 (2007).

[19] R. Das, T. Mills, L. Kwok, G. Maskel, I. Millett, S. Doniach, K. Finkelstein, D. Herschlag, and L. Pollack, "Counterion distribution around DNA probed by solution X-ray scattering", Phys. Rev. Lett. **90**, 188103 (2003).

[20] V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, "Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory", Biophys. J. **93**, 3202–3209 (2007).

[21] S. Kirmizialtin, A. R. Silalahi, R. Elber, and M. O. Fenley, "The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations", Biophys. J. **102**, 829–838 (2012).

[22] W. H. Braunlin, "NMR studies of cation-binding environments on nucleic acids", Advances in Biophysical Chemistry **5**, 89–140 (1996).

[23] S. L. Chang, S. H. Chen, R. L. Rill, and J. Lin, "Measurements of monovalent and divalent counterion distributions around persistence length DNA fragments in solution", J. Phys. Chem. **94**, 8025–8028 (1990).

[24] I. Morfin, F. Horkay, P. J. Basser, F. Bley, A.-M. Hecht, C. Rochas, and E. Geissler, "Adsorption of divalent cations on DNA", Biophys. J. **87**, 2897–2904 (2004).

[25] S. A. Pabit, K. D. Finkelstein, and L. Pollack, "Using anomalous small angle X-ray scattering to probe the ion atmosphere around nucleic acids", Methods Enzymol. **469**, 391–410 (2009).

[26] S. Kirmizialtin, S. A. Pabit, S. P. Meisburger, L. Pollack, and R. Elber, "RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations", Biophys. J. **102**, 819–828 (2012).

[27] K. Andresen, R. Das, H. Y. Park, H. Smith, L. W. Kwok, J. S. Lamb, E. Kirkland, D. Herschlag, K. Finkelstein, and L. Pollack, "Spatial distribution of competing ions around DNA in solution", Phys. Rev. Lett. **93**, 248103 (2004).

[28] M. Greenfeld and D. Herschlag, "Probing nucleic acid–ion interactions with buffer exchange-atomic emission spectroscopy", Methods Enzymol. **469**, 375–389 (2009).

[29] U. P. Strauss, C. Helfgott, and H. Pink, "Interactions of polyelectrolytes with simple electrolytes. II. Donnan equilibria obtained with DNA in solutions of 1-1 electrolytes", J. Phys. Chem. **71**, 2550–2556 (1967).

[30] D. Grilley, A. M. Soto, and D. E. Draper, "Direct quantitation of Mg^{2+}-RNA interactions by use of a fluorescent dye", Methods Enzymol. **455**, 71–94 (2009).

[31] S. A. Pabit, S. P. Meisburger, L. Li, J. M. Blose, C. D. Jones, and L. Pollack, "Counting ions around DNA with anomalous small-angle X-ray scattering", J. Am. Chem. Soc. **132**, 16334–16336 (2010).

[32] J. M. Blose, S. A. Pabit, S. P. Meisburger, L. Li, C. D. Jones, and L. Pollack, "Effects of a protecting osmolyte on the ion atmosphere surrounding DNA duplexes", Biochemistry **50**, 8540–8547 (2011).

[33] S. P. Meisburger, J. L. Sutton, H. Chen, S. A. Pabit, S. Kirmizialtin, R. Elber, and L. Pollack, "Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: beyond the wormlike chain model", Biopolymers **99**, 1032–1045 (2013).

[34] S. Kuwamoto, S. Akiyama, and T. Fujisawa, "Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants", J. Synchrotron Radiat. **22**, 462–468 (2004).

[35] V. Luzzati, F. Masson, A. Mathis, and P. Saludjian, "Étude, par diffusion centrale des rayons X, des polyelectrolytes rigides en solution. Cas des sels de Li, Na et Cs du DNA", Biopolymers **5**, 491–508 (1967).

[36] G. M. Giambaşu, T. Luchko, D. Herschlag, D. M. York, and D. A. Case, "Ion counting from explicit-solvent simulations and 3D-RISM", Biophys. J. **106**, 883–894 (2014).

[37] J. P. Bardhan, "Biomolecular electrostatics–I want your solvation (model)", Comput. Sci. Discovery **5**, 013001 (2012).

[38] Y. Marcus, *Ion solvation* (Wiley, 1985).

[39] Y. Marcus, "Ionic radii in aqueous solutions", Chem. Rev. **88**, 1475–1498 (1988).

[40] L. Pollack, "SAXS studies of ion-nucleic acid interactions", Annu. Rev. Biophys. **40**, 225–242 (2011).

[41] S. Hansen, "Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering", J. Appl. Crystallogr. **33**, 1415–1421 (2000).

[42] C. F. Anderson and M. T. Record Jr, "Ion distributions around DNA and other cylindrical polyions: theoretical descriptions and physical implications", Annu. Rev. Biophys. Biophys. Chem. **19**, 423–463 (1990).

[43] D. Svergun, C. Barberato, and M. H. J. Koch, "CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates", J. Appl. Crystallogr. **28**, 768–773 (1995).

[44] T. J. Macke and D. A. Case, "Modeling unusual nucleic acid structures", in *Molecular modeling of nucleic acids*, edited by N. B. Leontis and J. SantaLucia (American Chemical Society, Washington, DC, 1998), pp. 379–393.

[45] S. Kirmizialtin and R. Elber, "Computational exploration of mobile ion distributions around RNA duplex", J. Phys. Chem. B **114**, 8207–8220 (2010).

[46] S. A. Pabit, X. Qiu, J. S. Lamb, L. Li, S. P. Meisburger, and L. Pollack, "Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA", Nucleic Acids Res. **37**, 3887–3896 (2009).

[47] K. Andresen, X. Qiu, S. A. Pabit, J. S. Lamb, H. Y. Park, L. W. Kwok, and L. Pollack, "Mono-and trivalent ions around DNA: a small-angle scattering study of competition and interactions", Biophys. J. **95**, 287–295 (2008).

[48] L. Li, S. A. Pabit, S. P. Meisburger, and L. Pollack, "Double-stranded RNA resists condensation", Phys. Rev. Lett. **106**, 108101 (2011).

[49] C. R. Cantor, M. M. Warshaw, and H. Shapiro, "Oligonucleotide interactions. III. circular dichroism studies of the conformation of deoxyoligonucleolides", Biopolymers **9**, 1059–1077 (1970).

[50] A. V. Tataurov, Y. You, and R. Owczarzy, "Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids", Biophys. Chem. **133**, 66–70 (2008).

[51] S. S. Nielsen, M. Møller, and R. E. Gillilan, "High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell", J. Appl. Crystallogr. **45**, 213–223 (2012).

[52] T. C. Huang, H. Toraya, T. N. Blanton, and Y. Wu, "X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard", Journal of Applied Crystallography **26**, 180–184 (1993).

[53] D. Orthaber, A. Bergmann, and O. Glatter, "SAXS experiments on absolute scale with Kratky systems using water as a secondary standard", J. Appl. Crystallogr. **33**, 218–225 (2000).

[54] G. S. Kell, "Isothermal compressibility of liquid water at 1 atm.", J. Chem. Eng. Data **15**, 119–122 (1970).

[55] H. Durchschlag and P. Zipper, "Modeling the hydration of proteins: prediction of structural and hydrodynamic parameters from X-ray diffraction and scattering data", Eur. Biophys. J. **32**, 487–502 (2003).

[56] F. Merzel and J. C. Smith, "SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins", Acta Crystallogr. D Biol. Crystallogr. **58**, 242–249 (2002).

[57] S. Yang, M. Parisien, F. Major, and B. Roux, "RNA structure determination using SAXS data", J. Phys. Chem. B **114**, 10039–10048 (2010).

CHAPTER 2

POLYELECTROLYTE PROPERTIES OF SINGLE STRANDED DNA MEASURED USING SAXS AND SINGLE MOLECULE FRET: BEYOND THE WORMLIKE CHAIN MODEL

Steve P. Meisburger[1], Julie L. Sutton[1], Huimin Chen[1], Suzette A. Pabit[1], Serdal Kirmizialtin[2], Ron Elber[2], and Lois Pollack[1]

1. Applied and Engineering Physics, Cornell University, Ithaca, New York
2. Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas, USA.

Biopolymers **99**, 1032-45 (2013).

Abstract

Nucleic acids are highly charged polyelectrolytes that interact strongly with salt ions. Rigid, base-paired regions are successfully described with worm like chain models, but non base-paired single stranded regions have fundamentally different polymer properties because of their greater flexibility. Recently, attention has turned to single stranded nucleic acids due to the growing recognition of their biological importance, as well as the availability of sophisticated experimental techniques sensitive to the conformation of individual molecules. We investigate polyelectrolyte properties of poly(dT), an important and widely studied model system for flexible single stranded nucleic acids, in physiologically important mixed mono- and di-valent salt. We report measurements of the form factor and interparticle interactions using SAXS, end to end distances using smFRET, and number of excess ions using ASAXS. We present a coarse-grained model that accounts for flexibility, excluded volume, and electrostatic interactions in

these systems. Predictions of the model are validated against experiment. We also discuss the state of all-atom, explicit solvent Molecular Dynamics simulations of poly(dT), the next step in understanding the complexities of ion interactions with these highly charged and flexible polymers.

2.1 Introduction

The growing appreciation for the roles that nucleic acids play in biology calls for a thorough description of these biopolymers, including an understanding of how their mechanical properties couple to their biological function. Much of the effort thus far has focused on double stranded structures, which are well described by wormlike chain (WLC) models with ionic strength dependent persistence lengths that exceed 100 base pairs [1]. However, experience with other biopolymers, like proteins, demonstrates that although rigid structures are most amenable to experimental characterization, the flexible regions often impart biological function [2]. The most flexible regions of nucleic acids are non-base paired and include single stranded DNA (ssDNA) and RNA (ssRNA) regions that are involved in crucial biological processes. For example, polymerases unwind dsDNA, yielding stretches of ssDNA whose genetic information is transcribed into messenger ssRNA. The non-base-paired regions of ssRNA may be recognized by proteins involved in gene regulation or transport. The mechanical properties of ssRNA are exploited by riboswitches, where single stranded regions serve as actuators [3]. Finally, ssDNA is readily exploited in bioengineering, for example as tunable ligands for building nanoparticle superlattices [4].

Although the WLC model (and associated polyelectrolyte theory) has been successful in describing dsDNA, biophysical studies of single stranded nucleic acids in the last decade have found varying degrees of success applying WLC models. Estimates of the persistence lengths

and contour lengths in ssDNA and ssRNA vary widely among different experimental techniques, which have included fluorescence-based measurements [5–8], single molecule force extension [9–11], and small angle x-ray scattering (SAXS) [8, 12]. The polyelectrolyte theory describing electrostatic effects on polymer flexibility predicts an electrostatic component of the persistence length that has a power law dependence on the Debye screening length, where the exponent is different depending on assumptions about the intrinsic flexibility [13]. Two limiting theories, that of Odijk, Skolnick, and Fixman (OSF) [14, 15] and Barrat and Joanny (BJ) [16], predict exponents of 2 and 1, respectively. Experiments on dsDNA agree with OSF [1], but there is no consensus on whether the many experiments testing ssRNA and ssDNA fit either theory (reviewed in [12]).

To this end, we recently performed both SAXS and single molecule Förster Resonance Energy Transfer (smFRET) measurements of homopolymeric deoxythymidylate (poly(dT)) and uridylate (poly(rU)) molecules in solution and constrained a WLC model to simultaneously fit both the end-to-end distance measured by FRET and the entire scattering profile measured by SAXS [8]. SmFRET measurements over a wide range of monovalent and divalent salt concentrations were interpreted in this context. Surprisingly, we found that the power-law dependence of persistence length predicted by polyelectrolyte theory did not apply over the entire range of salt concentration. Furthermore the power law exponents fall between the OSF and BJ values and vary depending on the sugar moiety (ribose vs. deoxyribose) and the identity of the counterion (Mg^{2+} vs Na^{+}). However, the smFRET data also hinted at a possible reason for the discrepancy: divalent ions show an anomalously strong effect on structure, which suggests that ion condensation plays a major role in the conformations of ssDNA. The theories of OSF and BJ, which are based on the Debye-Hückel approximation for electrostatics, do not account for this phenomenon.

Another shortcoming of the WLC model is its neglect of excluded volume interactions. For

dsDNA, the diameter ($\sim$ 20 Å) is much smaller than the persistence length ($\sim$ 500 Å), but for single stranded nucleic acids they are comparable ($\sim 5-10$ Å). Thus, excluded volume might be safely neglected for dsDNA, but for ssDNA it is potentially important, even for short chains. Indeed, excluded volume effects have been observed in single molecule diffusion [7] and SAXS [12] studies of 8-128 nucleotide poly(dT), and in single molecule force spectroscopy measurements [10] of comparatively long ($\sim$ 10 kb) ssDNA. The picture of ssDNA as an electrostatically swollen coil, rather than a WLC, more strongly resembles recent models for highly charged intrinsically disordered proteins (IDPs) that invoke Flory theory [17], where excluded volume and solvent character play a dominant role. In one study, FRET measurements of IDP dimensions in an ionic denaturant were modeled analytically with chain monomers possessing a salt-dependent effective volume [18]. These results were consistent with a computational study using implicit solvent [19] that predicted expansion behavior of IDPs in accordance with hydrodynamic measurements. The success these approaches for charged IDPs suggests that similar models might be used for ssDNA.

In order to develop and test such a model, we carried out SAXS, single molecule FRET, and anomalous SAXS (ASAXS) measurements of 30 nucleotide poly(dT) (dT_{30}) in physiologically important mixed-salt solutions, where mono- and di-valent ions compete. We begin with a coarse-grained representation of ssDNA as a freely rotating chain of virtual bonds between the backbone C4' and P atoms [20, 21], and add hard-core excluded volume interactions between non-bonded monomers and mean-field electrostatics. The precise nature of ssDNA-ion interactions is unknown, and thus our model includes a single parameter for charge renormalization by strongly condensed ions. We determine this parameter for each salt condition by measuring the inter-molecular pair potential of dT_{30} molecules in solution using SAXS [22]. With our model constrained by the renormalized charge, we compute ensemble properties such as the radius of gyration, form factor, end-to-end distances, and ion distributions. These predictions are compared with corresponding measurements.

Like the WLC model and polyelectrolyte theory, our model involves coarse-graining and mean-field approximations. Ideally, single stranded nucleic acids and associated salt ions could be modeled without such compromises using explicit solvent molecular dynamics (MD). Recent MD predictions of the complex ion atmosphere around dsRNA have been directly compared with SAXS and ASAXS data [23, 24]. However, the large number of polymer degrees of freedom for single stranded nucleic acids poses a computational challenge that currently limits the technique to relatively short chains. Therefore, we also report explicit solvent MD simulations of dT_{10} with monovalent ions, and discuss the future applications of molecular dynamics in conjunction with ion-selective experimental techniques (such as ASAXS) for obtaining a more complete understanding of the interactions between ions and single stranded nucleic acids. Progress in this area will enable accurate modeling of folding and dynamics of these biologically important macromolecules.

2.2 Results

2.2.1 SAXS and smFRET measurements of dT_{30} compaction and charge screening in mixed Na^+ and Mg^{2+} salt

SAXS measurements were performed on dT_{30} in 20 mM NaCl and $MgCl_2$ concentrations of 0-20 mM, as described in Methods. For each solution condition, a series of DNA concentrations was measured to investigate the interparticle interference effect, where variation in the shape of the SAXS curve with concentration arises because of inter-particle interactions. Measurement of the interparticle interference serves a dual purpose: first for finding the form factor of dT_{30} by extrapolation to infinite dilution, and second for quantifying the strength of interactions between molecules [25]. Concentration-normalized SAXS profiles are plotted in Fig. 2.1. In 20

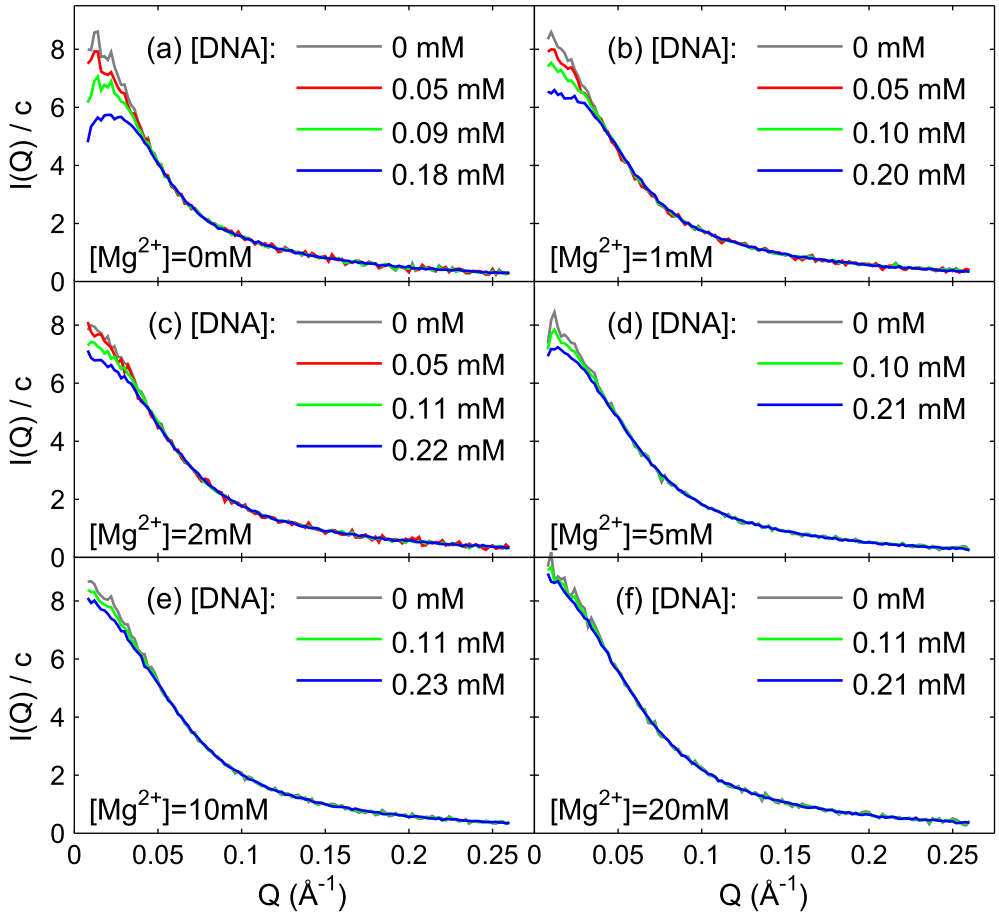

Figure 2.1: Concentration-normalized SAXS profiles of dT_{30} in 20 mM NaCl with added $MgCl_2$ and extrapolated form factors ([DNA]: 0 mM) in gray.

mM NaCl (Fig. 1a), the data show strong variation with DNA concentration for $q < 0.05\,\text{Å}^{-1}$, consistent with repulsive interactions between molecules [26]. As Mg^{2+} content increases (Fig. 2.1b-f), the net interparticle interaction weakens significantly but remains repulsive up to 20 mM $MgCl_2$.

The interparticle interference effect was quantified using an analysis of the SAXS data in terms of pair-wise interactions. In dilute solutions where the interference effect is small, the modulation of the scattering profile is linear in concentration and proportional to the second virial coefficient, B_2. This model was fit to the data to obtain B_2 for each solution condition, as described in Methods. The results are plotted in Fig 2.2. A sharp decrease is observed between

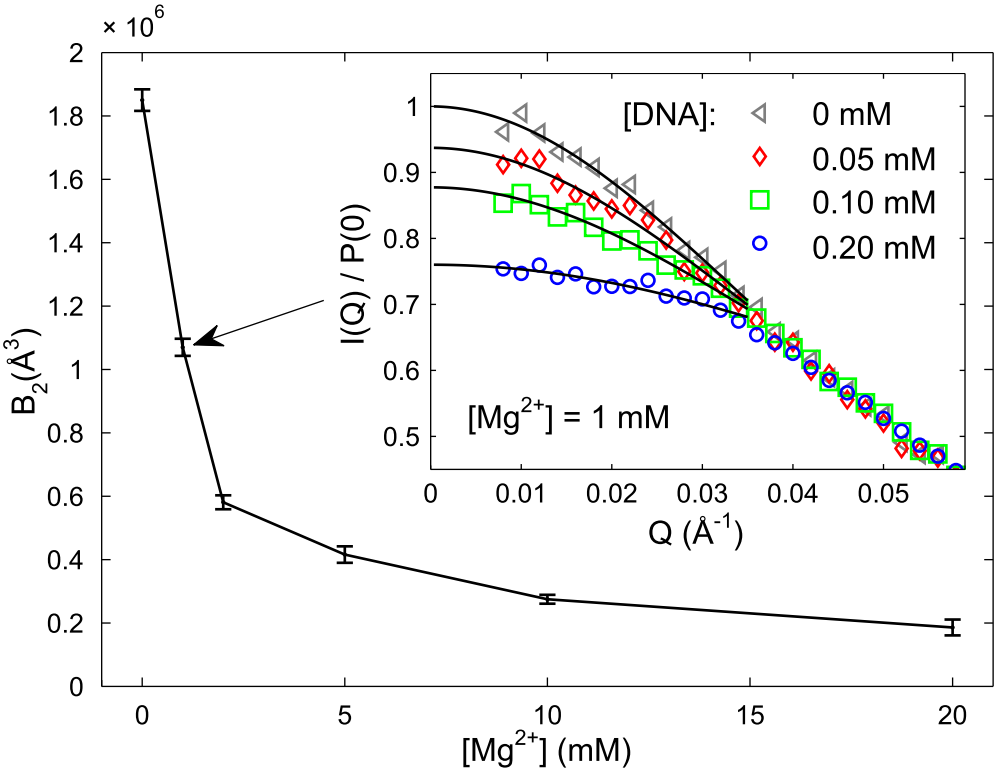

Figure 2.2: The second virial coefficient, B_2, was obtained by double-extrapolation of concentration-normalized SAXS profiles in Fig. 2.1 to $q = 0$ and $c = 0$, as described in Methods. The inset shows this extrapolation for dT_{30} in 20 mM NaCl and 1 mM $MgCl_2$.

0 and 2 mM $MgCl_2$, followed by a steady decline between 2 and 20 mM $MgCl_2$.

Next, the extrapolated form factors were compared to investigate the effect of Mg^{2+} on the chain conformations. The changes appear subtle in Fig. 2.1, but can be seen more clearly in a real-space analysis. The radius of gyration R_g and maximum dimension $D_{\max}$ were computed using a Bayesian indirect Fourier transform [27] implemented in MATLAB, and are given in Table 2.1. As can be seen from plots of R_g and $D_{\max}$ against Mg^{2+} concentration in Fig. 2.3a, the overall chain dimensions decrease as Mg^{2+} is added.

As a complementary probe of chain dimensions, smFRET measurements were performed on end-labeled dT_{30} in 20 mM NaCl with added $MgCl_2$, and the FRET-averaged end-to-end distance, $\langle R \rangle_{\mathrm{FRET}}$, was determined from the FRET efficiency and the Förster radius as described

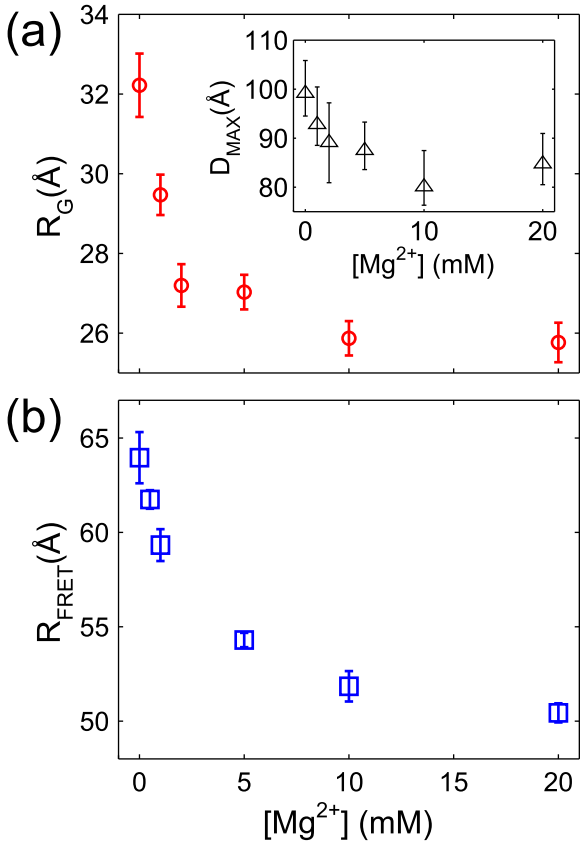

Figure 2.3: Chain dimensions of dT_{30} measured by SAXS and smFRET in 20 mM NaCl with added $MgCl_2$. (a) The radius of gyration R_g and the maximum dimension $D_{\max}$ were determined from SAXS data by Bayesian indirect Fourier transform of the extrapolated form factors shown in Fig. 2.1. Error bars correspond to the uncertainty of the fit. (b) The FRET-averaged end-to-end distance, $\langle R \rangle_{\mathrm{FRET}}$, was measured with smFRET of freely diffusing, fluorescently labeled dT_{30} in buffer containing 20 mM NaCl and 0, 0.5, 1, 5, 10, and 20 mM $MgCl_2$. Error bars show the standard deviation of at least four repeated measurements.

[Mg^{2+}] (mM)	R_{FRET}(Å)	R_g(Å)	D_{MAX}(Å)	$B_2(\times 10^6 Å^3)$
0.0	64.0 ± 1.4	32.2 ± 0.8	99 ± 6	1.850 ± 0.034
0.5	61.7 ± 0.5	n.d.	n.d.	n.d.
1.0	59.3 ± 0.8	29.5 ± 0.5	93 ± 6	1.070 ± 0.027
2.0	n.d.	27.2 ± 0.5	89 ± 8	0.581 ± 0.022
5.0	54.3 ± 0.4	27.0 ± 0.4	88 ± 5	0.416 ± 0.026
10.0	51.8 ± 0.8	25.9 ± 0.4	80 ± 6	0.275 ± 0.014
20.0	50.4 ± 0.5	25.8 ± 0.5	85 ± 5	0.186 ± 0.025

Table 2.1: Experimental data in Figs. 2.2 and 2.3, for dT_{30} with 20 mM NaCl and varying $MgCl_2$ (n.d. = not determined).

in Methods. The results are given in Table 2.1, and plotted as a function of Mg^{2+} concentration in Fig. 2.3b. An increase in Mg^{2+} concentration is accompanied by a decrease in $\langle R \rangle_{\mathrm{FRET}}$. As was observed with SAXS, Mg^{2+} has the effect of decreasing overall chain dimensions.

2.2.2 The number of excess monovalent ions around dT_{30}

The number of monovalent ions around dT_{30} in excess of the bulk concentration was measured using two different SAXS techniques: heavy ion replacement [28] and ASAXS [29]. While these techniques have been applied double-stranded RNA and DNA, ion populations for single stranded nucleic acids have not been previously reported. Both techniques obtain information about the ion atmosphere by varying the scattering contrast of the ions, either by comparing high-Z and low-Z atomic species (heavy ion method), or by exploiting the energy-dependent scattering factor for the ion of interest near an x-ray absorption edge (ASAXS). This information can include the number of ions per macromolecule when the scattering patterns are placed on an absolute scale. For the heavy ion method, we compare the scattering from two identicaly-prepared solutions of 0.1 mM dT_{30} in 100 mM RbCl or 100mM NaCl. ASAXS measurements were performed on dT_{30} with 100 RbCl at two energies near the K-edge of Rb, as described in methods.

For the heavy ion experiments, the macroscopic scattering cross-section, $d\Sigma(0)/d\Omega$, depends on the contrast of the ions at $q = 0$ as,

$$\begin{aligned} \frac{d\Sigma(0)}{d\Omega} &= n r_0^2 \left(\Delta N^e_{\mathrm{total}}\right)^2 \\ &= n r_0^2 \left(\Delta N^e_{DNA} + N_{Ions} \times \Delta N^e_{Ion}\right)^2 \end{aligned} \tag{2.1}$$

where r_0 is the classical electron radius, n is the concentration (N/V) of DNA molecules, and

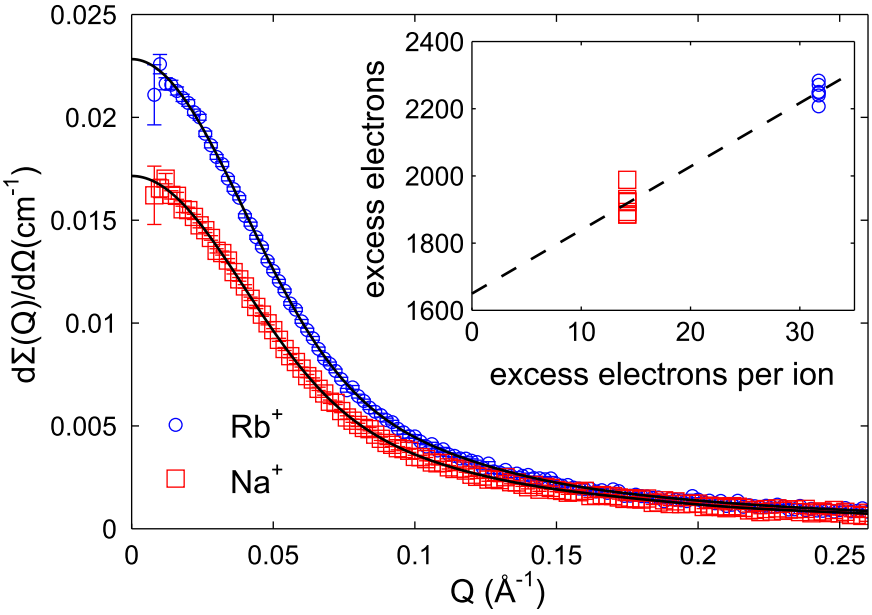

Figure 2.4: Quantitative heavy atom replacement measurement of the ion atmosphere around dT_{30}. The macroscopic scattering cross section of dT_{30} increases when electron-rich Rb^+ ions comprise the ion atmosphere, as opposed to Na^+. The number of excess monovalent ions was determined as shown in the inset, from a plot of total scattering contrast vs. the contrast per ion.

$\Delta N^e_{\text{total}}$ is the total excess electron density contrast, with contributions from DNA and ion components: ΔN^e_{DNA} is the number of excess electrons for DNA, and ΔN^e_{Ion} is the number of excess electrons per ion, which we calculate using

$$\Delta N^e_{Ion} = N^e_{Ion} - \rho_w V$$

where ρ_w is the electron density of water and V is the absolute limiting partial molar volume of the ion ($V_{Na} = -12.6\,\text{Å}^3$ and $V_{Rb} = 12.8\,\text{Å}^3$ [30]). The number of excess ions around dT_{30} was determined from a plot of ΔN^e_{total} vs. ΔN^e_{Ion}, shown in Fig. 2.4. The slope yields $N_{Ions} = 18.9 \pm 1.1$.

The ASAXS technique for measuring ion numbers takes advantage of the anomalous scattering properties of Rb ions [29]. The real part of the scattering factor of Rb, f'_{Rb}, changes by several electrons as the x-ray energy is varied below the K-edge at 15.2 keV. Calibrated SAXS

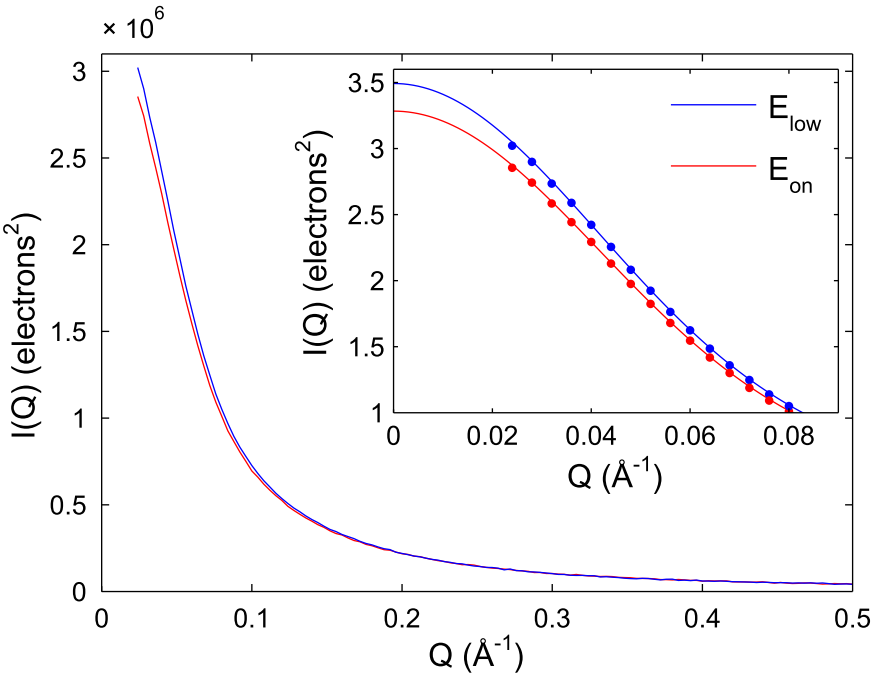

Figure 2.5: Quantitative ASAXS measurements to probe the monovalent ion atmosphere around dT_{30}. The scattering contrast of Rb^+ ions around dT_{30} was varied by tuning the X-ray energy between E_{on} and E_{lo}, defined in the text. The number of ions was determined by the change in the forward scattering with extrapolation to $q = 0$, as shown in the inset.

profiles from a solution of dT_{30} in 100 mM RbCl are plotted in Fig. 2.5, and show a decrease in forward scattering near the Rb edge due to the lower contrast of Rb at that energy. The number of Rb ions was calculated as described in Methods. With ASAXS, we find 19.1 ± 0.9 excess Rb ions around dT_{30}.

2.2.3 A charged, freely rotating chain model for ssDNA

DNA molecules have six bonds per nucleotide that can take on many orientations. For efficient computational sampling of conformations, it is desirable to reduce the number of degrees of freedom by coarse-graining. The WLC model is one way of achieving this, but its continuously deformable nature makes it a more useful analytic tool than a computational one. WLCs are members of a family of inextensible statistical chains that include freely-rotating and freely-

jointed chains (FRC and FJC, respectively) that are statistically equivalent in limit of large number of segments (or with contour length much longer than persistence length). Therefore, an FRC model for ssDNA is developed in analogy to the virtual bond description of polynucleotide backbones, which reduces the number of bonds per nucleotide to two. For simplicity, the virtual bond segments are assigned equal length l_0 and valence angle θ, while the dihedral angles are free to rotate. The synthetic poly(dT) molecules modeled here lack 5' phosphate groups, so a chain with N bases has $N-1$ phosphates and $2(N-1)$ virtual bonds. Allowable conformations are restricted by a hard-core excluded volume interaction, such that non-bonded atoms have a minimum allowable separation, d.

Polyelectrolytes such as ssDNA swell in response to the ionic environments. In order to model electrostatic effects, charges (with valence $Z=-1$) are placed at the P-coordinates along the chain. The parameter a is the closest approach for a hydrated ion and (virtual) chain atom, defining the boundary between the macromolecule and the electrolyte. The geometric parameters l_0, θ, a, and d are illustrated in Fig. 2.6.

Electrostatic effects are modeled using the Poisson-Boltzmann (PB) equation as a starting point. Outside the macromolecule, where the solvent has a uniform dielectric constant $\varepsilon = 78.5$ and an electrolyte with m species of ion with valence $z_i$and bulk density n_i^{∞}, the electrostatic potential satisfies the PB equation,

$$\nabla^2\Phi(\vec{r}) = -\frac{1}{\varepsilon\varepsilon_0}\sum_{i=1}^{m} e_0 z_i n_i^{\infty} \exp(-z_i e_0 \Phi(\vec{r})/k_B T) \tag{2.2}$$

where ε_0 is the permittivity of vacuum, and e_0 is the elementary charge. The linearized form of PB, also called the Debye-Hückel (DH) equation, is

$$\nabla^2\Phi(\vec{r}) = \kappa^2\Phi(\vec{r}) \tag{2.3}$$

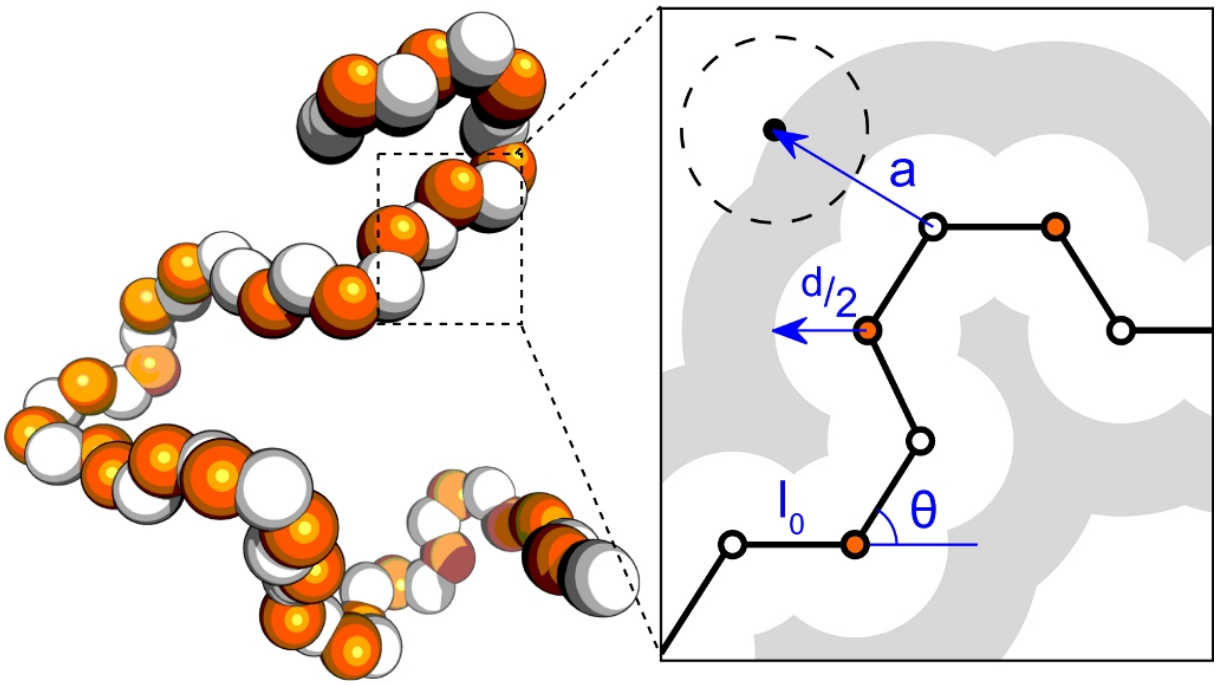

Figure 2.6: Illustration of geometric parameters for the charged, freely rotating chain model of ssDNA. Virtual bonds of length l_0 span backbone C4' and P atoms (open and orange filled circles, respectively). The valence angle between bonds is θ and d is the chain diameter. For electrostatics calculations, each P position is assigned a negative charge and the distance a defines ion accessibility. A representative ion is drawn as a dashed circle, and the ion-inaccessible region is shaded in gray. The cartoon drawing on the left shows a virtual bond model for dT_{30} rendered using Pymol version 1.2r1 (DeLano Scientific LLC).

where $\kappa = (8\pi l_B A_0 I)^{1/2}$ is the inverse Debye length with $A_0 = 6.022 \times 10^{-4} \text{Å}^{-3} M^{-1}$, $l_B = e_0^2/(4\pi\varepsilon_0\varepsilon k_B T)$ is the Bjerrum length, equal to 7.14 Å for water at room temperature, and I is the ionic strength.

The DH equation is an accurate approximation of PB when the potential is weak ($|\Phi| < k_B T/e_0$), far from the macromolecule. However, the magnitude of the solution to the DH equation can be inaccurate because it depends on the application of Gauss' law at the molecular surface, where the potential is in general much larger than $k_B T/e_0$. A common way to "rescue" the DH solution is to renormalize the charge of the macromolecule in the DH solution so that it approximates the true potential at large distances [31]. The charge renormalization factor, $f = Z_{eff}/Z$ can be derived by solving the PB and DH equations for a model system (e.g. [32])

or left as an experimentally determined parameter (e.g. [22]). Here, we choose the latter option.

Using the linearity of the DH equation, the electrostatic free energy for the chain is the sum over all pair-wise electrostatic interactions:

$$W_{chain} = f^2 \frac{1}{2} \sum_i \sum_{j \neq i} W_2(|\vec{r}_i - \vec{r}_j|) \tag{2.4}$$

where $W_2(r)$ is the free energy for a pair of phosphates at a distance r apart, and the summations run over all P coordinates in the chain. We approximate $W_2(r)$ using the repulsive DLVO potential [33] for charged spheres with diameter d:

$$\frac{W_{\mathrm{DLVO}}(r)}{k_B T} = l_B \left(\frac{Z}{1 + \kappa d/2} \right)^2 \frac{\exp(-\kappa(r-d))}{r} \tag{2.5}$$

2.2.4 Insight into model parameters from crystal structures of poly(dT)

In order to choose physically realistic geometric parameters for the model, we examined x-ray crystal structures deposited in the Protein Data Bank (PDB) that contain single-stranded poly(dT) bound to proteins. Continuous strands of poly(dT) resolved in four crystal structures (Table 2.2) were analyzed using the virtual bond representation. Histograms for the virtual bond length and valence angle are shown in Fig. 2.7a and b. Bond lengths l_0 ranged from 2.9 - 4 Å, while the valence angle θ spanned $10 - 110^\circ$. The average values $l_0 = 3.69$ and $\theta = 57.8^\circ$ were chosen for the model.

The minimum distance between two chain atoms, d, should be comparable to the Van der Waals diameter of the sugar phosphate backbone, in the range of roughly $5.4 - 6.0$ Å. The distance of closest approach for ions and DNA, a, is not straightforward to determine because of hydration effects. Recently, Molecular Dynamics simulations of ions around an RNA duplex

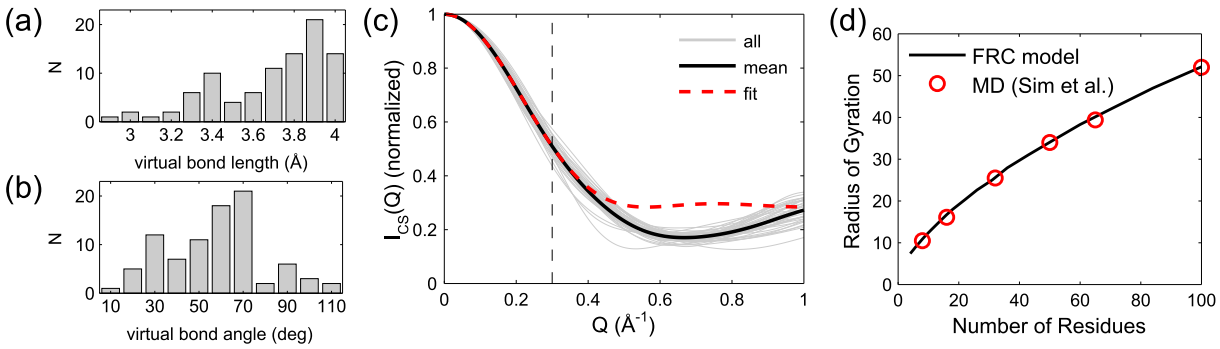

Figure 2.7: Virtual bond analysis of atomic models of poly(dT). (a,b) Histograms of bond lengths, l_0 and angles θ were generated from crystallographic models described in Table 2.2. The average virtual bond parameters are $l_0 = 3.69$ Å and $\theta = 57.8°$. (c) Effective scattering of the cross-section $I_{CS}(q)$ was calculated from four-nucleotide subsets of the atomic models as described in the text. The solid line shows the average, and the dashed line shows a convenient fit in the experimental q-range ($0 < q < 0.3\,\text{Å}^{-1}$) (see text). (d) The radius of gyration was computed for virtual bond chains with varying numbers of residues using the average parameters from (a) and (b), and a chain diameter $d = 5.6$ Å. The results (solid line) are plotted with "sterics only" molecular dynamics simulation results (open circles) reported in Ref. [12].

PPB ID	Chain	Residues	Resolution	Structural Context
3VDY	F	1-10	2.8 A	B. Subtilis, ssDNA-binding protein B [34]
2VW9	C	18-26	2.3 A	H. Pylori, ssDNA-binding protein [35]
4GOP	K	1-25	3.1 A	Eukaryotic Replication protein A [36]
1XHZ	G	1-5	2.7 A	Phi29 DNA polymerase [37]

Table 2.2: Poly(dT) models from the Protein Data Bank (PDB) analyzed here.

with explicit solvent [23] revealed two binding modes for Na^+ characterized by ion-O2P distances of 2.3 and 4.5 Å, while hydrated Mg^{2+} approaches O2P at an intermediate distance of ~ 4 Å. Adding 1.5 Å for P-O2P bond, the distance a should fall in the range of $3.8 - 6.0$ Å. For simplicity, ions and chain atoms were treated equivalently, with the choice of $a = d = 5.6$ Å.

In order to calculate accurate SAXS profiles, particularly at high-q, it is important to account for finite thickness of the chain. As in previous SAXS studies of poly(dT), the x-ray scattering pattern is calculated using

$$I(q) = P_{\text{chain}}(q) \times I_{CS}(q) \tag{2.6}$$

where $I_{CS}(q)$ accounts for the finite thickness. In earlier measurements of poly(dT) [8], a cylinder model was used for I_{CS}(q)

$$I_{CYL}(q,R) = (2J_1(qR)/(qR))^2 \tag{2.7}$$

with the radius R as a fitting parameter. Here, we improve on this method by calculating an average $I_{CS}(q)$ directly from the crystal structures listed in Table 2.2. The chains were divided into 4-nucleotide segments, and the scattering patterns were computed using CRYSOL [38] with default hydration parameters. Then, $I_{CS}(q)$ was found by dividing the scattering profile by $P_{\text{chain}}(q)$ calculated from the virtual bond representation of the same structure, as described in Methods. The profiles and their average are shown in Fig. 2.7c. The average $I_{CS}(q)$ resembles $I_{CYL}(q,R)$ used previously, however the fit can be made essentially exact for $q < 0.3\text{Å}^{-1}$ by adding a constant: $I_{CS}(q) \approx (1-c) \times I_{CYL}(q,R) + c$ with $R = 6.812\text{Å}$ and $c = 0.2838$.

To verify that the above choice of model parameters gives a reasonable approximation of the intrinsic flexibility and excluded volume of poly(dT), the mean squared radius of gyration for ensembles of chains was compared with "sterics only" molecular dynamics calculations recently reported for poly(dT), as the number of nucleotides was varied from N = 8 to 100 [12], shown in Fig. 2.7d. Agreement between the coarse-grained and all-atom representations is excellent. Furthermore, we find that the radius of gyration follows a polymer scaling law [39], $R_g \propto N^{\nu}$ with $\nu = 0.62$, which matches the reported value from an earlier MD calculation of poly(dT) for N = 8 to 128 nucleotides [7].

2.2.5 Determination of the charge renormalization parameter from measurements of B_2

With the geometric parameters for the poly(dT) model fully specified, we turn to the electrostatic parameters. The Debye screening length κ^{-1} is determined solely by the ionic strength of the solution, but f must be fixed experimentally for each ionic condition by matching B_2. Therefore, for each experimental value of κ, and for a range of f values, B_2 was computed as described in Materials and Methods. Briefly, a set of chains, $\Omega_{\kappa,f}$ was generated using the Metropolis algorithm [40] with the energy function W_{chain} for every (κ, f) pair of parameters. Next, pairs of chains were drawn from $\Omega_{\kappa,f}$, and their centers of mass were displaced from each other by a random vector $\boldsymbol{R}$, chosen with uniform probability density within a sphere of radius R_{max}. If the chains clashed, the interaction potential was assigned $W_m = \infty$. Otherwise, W_m was found in an analogous way to the single chain energy in Equations 2.4-2.5.

After M iterations, the second virial coefficient was computed,

$$B_2 = \frac{1}{2} V_{excl} = \frac{V_{max}}{2M} \sum_{m=1}^{M} (1 - \exp(-W_m)) \tag{2.8}$$

where $V_{max} = (4/3)\pi R_{max}^3$. The value of R_{max} must be large enough that chains have negligible interactions at that distance. Therefore, B_2 was calculated with increasing values of R_{max} until no significant change was observed.

The result of the calculation is a set of curves for B_2 vs. f, shown in Fig. 2.8a. For each salt condition, experimental values of B_2 ($\pm$ standard errors) were mapped onto the parameter f. Results are plotted vs. Mg concentration in Fig. 2.8b and Table 2.3.

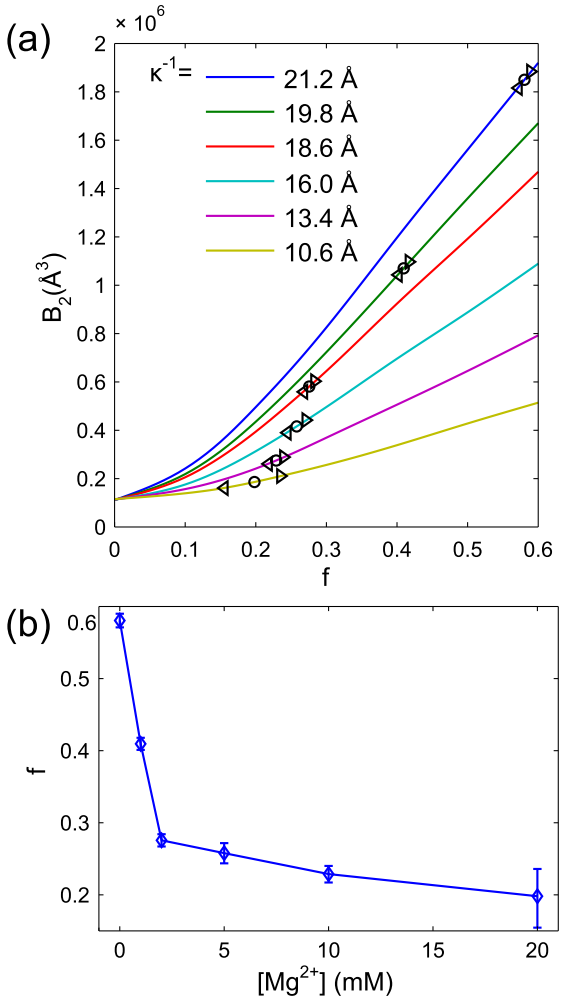

Figure 2.8: Relationship between the second virial coefficient B_2 and the charge renormalization parameter f for the FRC model of dT_{30}. (a) B_2 vs. f was calculated for values of the Debye screening lengthκ^{-1} corresponding to 20 mM NaCl plus 0, 1, 2, 5, 10, or 20 mM $MgCl_2$. Experimental values of B_2 (circles) $\pm$ standard errors (triangles) were mapped onto f at each salt condition. (b) Plot of f vs Mg^{2+} concentration obtained from (a).

$[Mg^{2+}]$ (mM)	κ^{-1}(Å)	f	$\langle R_{G0}^2 \rangle^{1/2}$(Å)	$\langle R_G^2 \rangle^{1/2}$(Å)	$\langle R^2 \rangle^{1/2}$(Å)	$\langle R \rangle_{\mathrm{FRET}}$(Å)
0	21.21	0.581	30.02	30.44	83.35	71.24
1	19.78	0.409	27.43	27.84	74.28	64.71
2	18.61	0.276	25.63	26.10	68.10	60.61
5	16.04	0.258	25.26	25.77	66.78	59.75
10	13.42	0.229	24.82	25.35	65.29	58.81
20	10.61	0.198	24.44	25.00	64.05	58.03

Table 2.3: Model parameters for dT_{30} and predicted ensemble average dimensions.

2.2.6 Prediction of chain conformations

With the value of f constrained by the measurement of B_2 at each salt condition, it is now possible to predict other observables using the model, such as the radius of gyration, the end to end distance distribution, the scattering profile, and the interparticle interference function. These predictions are compared directly with experimental data.

At each condition, a set of chain conformations was generated by Monte Carlo sampling. The end-to-end distance histogram was converted to $\langle R \rangle_{\mathrm{FRET}}$ as described previously [8]. The average pair distance distribution function P(r) was used to calculate the mean square radius of gyration of the chain,

$$\langle R_{g0}^2 \rangle = \frac{1}{2} \left(\sum_i \bar{h}_i \right)^{-1} \sum_i \bar{h}_i r_i^2 \tag{2.9}$$

and the form factor,

$$P_{\mathrm{chain}}(q) = \left(\sum_i \bar{h}_i \right)^{-1} \sum_i \bar{h}_i \frac{\sin(q\, r_i)}{q\, r_i} \tag{2.10}$$

To account for the finite thickness of the ssDNA in calculation of the scattering profiles, the form factor was multiplied by the effective scattering of the cross-section derived from crystal structures (see Fig. 2.7): $I(q) = P_{chain}(q) \times I_{CS}(q)$. The effect of multiplication by $I_{CS}(q)$ is that the SAXS radius of gyration is slightly larger than that of the skeletal chain, $\langle R_g^2 \rangle = \left\langle R_{g0}^2 \right\rangle + (5.04\text{Å})^2$, where 5.04 Å is the effective radius of gyration according to a Guinier fit of $I_{CS}(q)$ at low-q. Calculation of the interparticle interference function follows the method of Hubbard and Doniach [41] generalized for non-identical particles, and is described in Methods. Numerical results are given in Table 2.3.

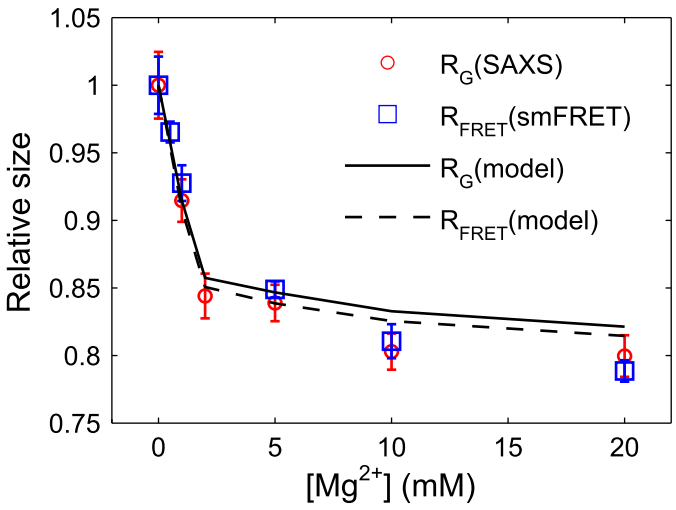

Figure 2.9: Chain compaction measurements by SAXS and smFRET are compared with model predictions. The ratio between R_g (or $\langle R \rangle_{\mathrm{FRET}}$) in 20 mM NaCl with added $MgCl_2$ and the value in 20 mM NaCl alone is plotted as a function of added Mg^{2+}.

2.2.7 Chain compaction: model vs. experiment

Comparison of modeled and experimental values for R_g and $\langle R \rangle_{\mathrm{FRET}}$ in Tables 2.1 and 2.3 shows that the SAXS predictions are about 4% lower than observation, while FRET predictions are about 10% higher than observed. However, on a relative scale, the model captures the extent of chain compaction. In Fig. 2.9, R_g and $\langle R \rangle_{\mathrm{FRET}}$ from the model and experiment were normalized by their value at 20 mM NaCl. The predictions reproduce the magnitude of chain compaction and overall shape of the curve at all Mg concentrations. Furthermore, the model suggests that $\langle R \rangle_{\mathrm{FRET}}$ and R_g are nearly proportional, and correspondingly the SAXS and smFRET measurements of the chain size are proportional within error.

2.2.8 Polymer scaling properties of poly(dT): model vs. experiment

At high-q, the polymer scaling behavior becomes important, and this is a valuable way to test whether the data are consistent with a swollen polymer. The inset of Fig. 2.10 compares the

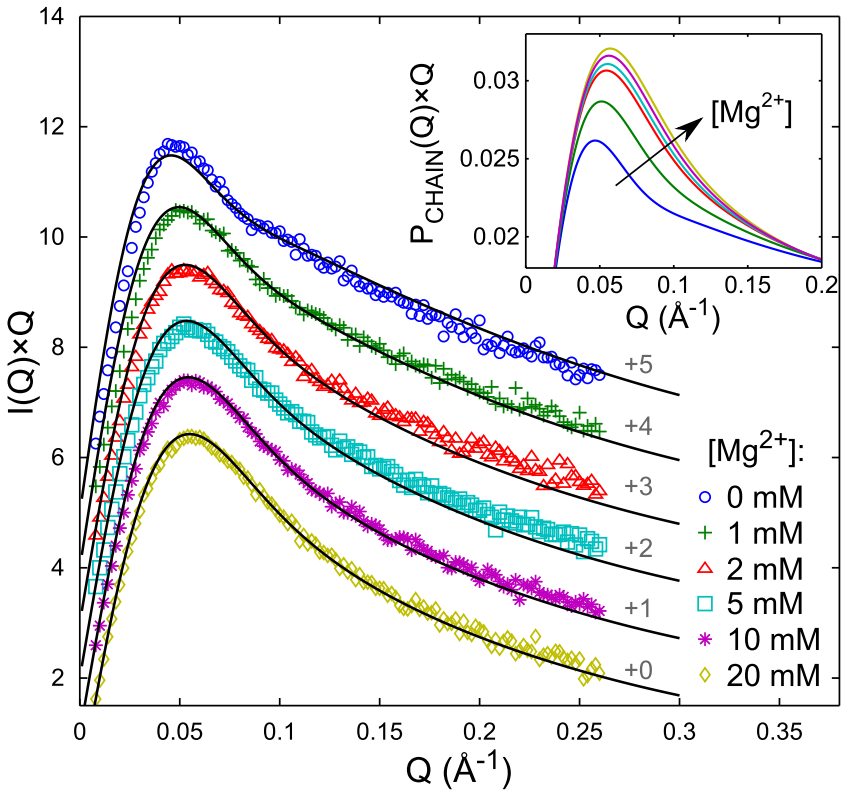

Figure 2.10: At high q, the form factor is sensitive to polymer scaling properties. Model form factors at each Mg concentration indicated in the legend and 20 mM NaCl are multiplied by q to emphasize the high-q region, and plotted in the inset. To compare with experiment, form factors were multiplied by the effective cross section and scaled to match experimental curves at $q = 0.1\,\text{Å}^{-1}$. The curves are offset vertically (for clarity) as indicated next to each curve.

calculated form factor of the chain multiplied by q, at each salt condition, where the arrow indicates the direction of increasing Mg^{2+} . In the main part of Fig. 2.10, the SAXS data are plotted as $I(q) \times q$ vs q, along with model predictions. The high concentration (0.2 mM) DNA data are shown, rather than the extrapolated form factor, because they have much higher signal-to-noise: interparticle interference is present at low-q, but $I(q)$ for $q > 0.05\text{Å}^{-1}$ can be compared with the predictions. The model captures the change in shape of the scattering profile as Mg is added, indicating that the interpretation of poly(dT) as a swollen coil is essentially correct.

The predicted and experimental interparticle interference functions plotted in Fig. 2.11

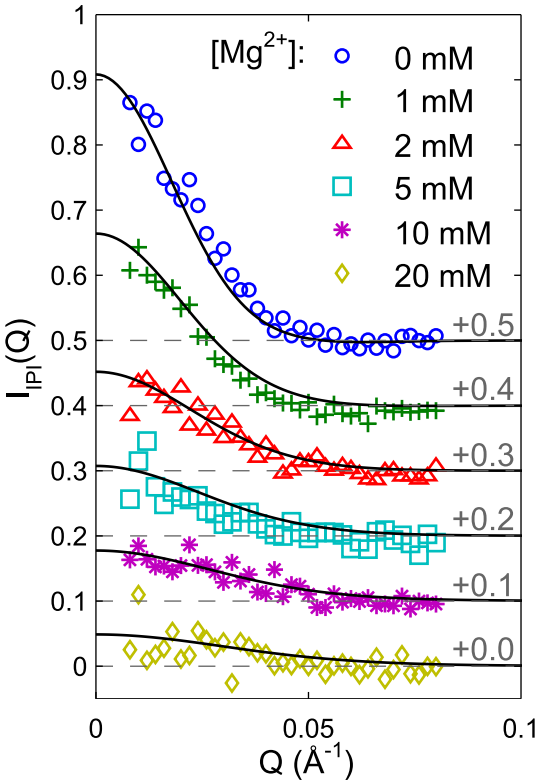

Figure 2.11: Interparticle interference functions $I_{\mathrm{IPI}}(q)$ for model (solid lines) and experiment (points) at the highest DNA concentration (nominally 0.2 mM) are plotted for each Mg^{2+} concentration, given in the legend, and offset vertically by the amount indicated.

show a change in shape consistent with repulsive interactions between chains becoming shorter-ranged with increasing Mg. Note that because B_2 was determined by experiment, the magnitude of the interparticle interference function must agree; however, the shape of the interference function is an independent prediction.

2.2.9 Ion atmosphere around dT_{30}

The charge renormalization parameter can be interpreted physically in terms of the ion atmosphere using PB theory, where $(1-f)$ corresponds to a fraction of the macromolecule's charge that is compensated by strongly condensed ions [31]. The physical interpretation of f becomes

more complicated in mixed salt, so here we concentrate on the case of 20 mM Na with no added Mg. Then, the number of strongly condensed sodium ions is

$$N_+^{\text{condensed}} = (1-f)\,N_{\text{phos}}$$

For dT_{30}, $(1-0.581)\times 29 \sim 12.1$ ions. The number of strongly condensed ions is not easily measured experimentally; rather, ASAXS and heavy-ion report the total number of excess counterions,

$$N_+^{\text{excess}} = N_+^{\text{condensed}} + N_+^{\text{diffuse}}$$

To predict N_{excess}^+, one needs to calculate the number of diffusively bound ions, N_{diffuse}^+. Unfortunately, there is no unique way to distinguish between a strongly condensed and diffusively bound ion within PB theory on the basis of distance from the macromolecule or electrostatic potential, except for simple geometries [31]. Here, we follow previous work (e.g. [42]) and define the diffuse ions as those bound with a potential less than k_BT/e_0. According to PB theory, the excess counterion density is

$$\Delta n_+(\vec{r}) = n_+^{\infty}(\exp(-e_0\,\Phi(\vec{r})/k_BT) - 1) \tag{2.11}$$

where n_+^{∞} is the bulk density. The potential is given by the renormalized DH solution for charged spheres with radius a:

$$\frac{e_0\,\Phi(\vec{r})}{k_BT} = f\sum_{i=1}^{N_{phos}}\left(\frac{Z\,l_B}{1+\kappa a}\right)\frac{\exp(-\kappa(|\vec{r}-\vec{r}_i|-a))}{|\vec{r}-\vec{r}_i|} \tag{2.12}$$

The total number of diffusively bound ions, N_{diffuse}^+ is the integral of over the electrolyte

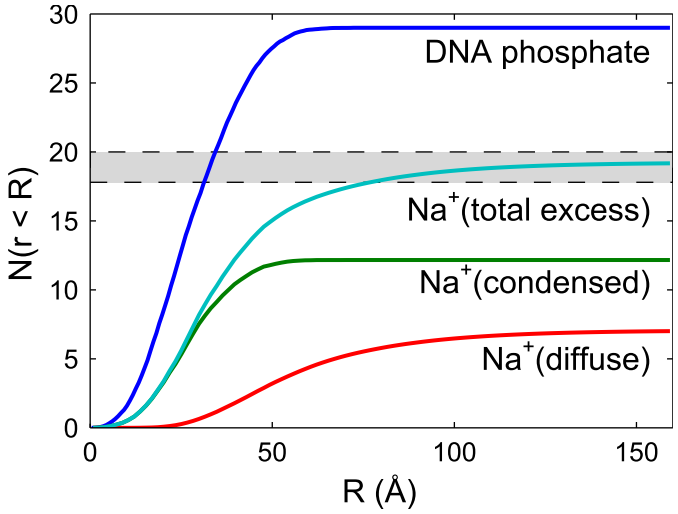

Figure 2.12: Number of charged phosphate groups and excess counterions within a radius R of the chain center of mass computed using the model for dT_{30} in 20 mM NaCl. The total number of excess sodium ions is the sum of the ions in diffuse and condensed regions, defined in the text. The experimental values for the number of excess monovalent ions are represented as a gray box bounded by dashed lines showing the $\pm 1\sigma$ confidence interval.

volume where $|\Phi(\vec{r})| < k_B T/e_0$. The calculation of N^+_{diffuse} for poly(dT) is slightly more complicated: the integral also must be performed over the distribution of chain conformations. To accomplish this efficiently, Monte Carlo integration was performed as described in Methods.

The calculated number of excess ions within a radius R of the center of mass of the chain is shown in Fig. 2.12. The distribution of condensed ions follows the phosphate distribution, while the diffuse ion cloud extends outside the chain. The number of excess Na^+ approaches a value of ~ 19 far from the chain. This calculation agrees with the measured values within experimental error.

2.2.10 Explicit solvent molecular dynamics simulations of dT_{10}

The number of excess counter ions was also computed by all atom Molecular Dynamics simulations (MD). MD accounts for ions and water as well as the atoms of the nucleic acid explicitly,

in contrast to PB approaches that model the solvent and ions as a continuous medium [31]. Here we report simulation results for dT_{10} in 50 mM Na^+, which yield the number of excess Na^+ directly (see Methods). The simulations find 5 ± 0.3 ions, around dT_{10}, corresponding to 56% charge compensation (polymer charge of 9) due to excess ions. The results are lower than those obtained from experiments described above, where 66% (19 ions and polymer charge of 29) of the charge is compensated by excess Na^+ ions around dT_{30}. The variance between experiment and MD simulation is discussed below.

2.3 Discussion

We applied a simple polymer model for poly(dT), in which ionic strength dependent interactions between chains are fixed by experimental measures of the second virial coefficient and observables are predicted without additional input from experiment. This represents an improvement over previous applications of WLC, where the ionic strength dependent persistence length was a fitting parameter. As a simplifying assumption, the bond lengths and angles are fixed, and the potentials determine the distribution of chain conformations. While it would seem that this is fundamentally different from the WLC model (with no explicit potential, and a salt-dependent persistence length), it is merely a diffent way accounting for bond correlations. The main difference is that for excluded volume, the bond correlations may be much longer-ranged than in the WLC, where by definition they decay exponentially.

A critical assumption of the proposed model is that the same driving force for inter-chain repulsion (e.g. screened electrostatics) is responsible for changes in the chain dimensions. Measurements by SAXS and smFRET allowed us to test this assumption. The model predictions reproduce the salt-dependent changes in chain dimensions, as well as the number of ions the chains attract.

While model predictions for the radius of gyration and FRET distance agree on a relative basis, they fall outside the experimental error on an absolute basis. Since for this study we focused on constraining the electrostatic aspects of poly(dT) rather than the backbone properties, it is possible that the geometrical parameters could be tuned for better agreement with experiment. However, we also neglected the scattering of the ion atmosphere when computing Rg, as well as the possible influence of dye labels and chain dynamics on FRET [43]. Future studies that rigorously test MD using FRET and SAXS data will account for these effects.

Interestingly, the MD simulations do not reproduce the same fraction of charge neutralization as measured experimentally. Although shorter chains and lower ambient salt concentrations are used for MD than for experiment (dT_{10} as opposed to dT_{30}, and 50 mM Na as opposed to 100 mM Na), we expect the number of excess ions to be similar across these different conditions [44]. The deviation may be statistical (even hundreds of nanoseconds may be insufficient to sample the relevant range of conformations of dT_{10}). Moreover the small number of ions in the simulation box (see Methods) implies other statistical inaccuracies. While the differences can also be a result of force field tuning, we comment that all atom simulations quantitatively captured ion distribution near the A form of dsRNA [24].

2.4 Summary and Outlook

The work described herein focuses on poly(dT) measured over a relatively small range of ionic conditions, less than, but approaching physiological ionic strength. Future work will extend these measurements to provide a broader test of the model. At lower ionic strength, repulsive forces increase dramatically enabling more robust comparison with electrostatic parameters. At much higher ionic strength, the possibility of achieving theta solvent conditions [10] will allow an assessment of the balance between attractive and repulsive forces. Opportunities also exist

to explore the distinctly different roles of mono and multi (e.g. di-) valent ions in providing charge compensation. Although we described measurements in mixed salt solutions, all of these important details were absorbed by the charge renormalization parameter, f. Our simple model does not address the distinctly different contributions of mono- or di- valent ions to screening. The surprisingly efficient charge compensation measured in the presence of Mg^{2+} in this simple system may be an ideal vehicle for probing effects such as fluctuations, correlations or polarizability. The inclusion of models of more complex electrostatic environments [31, 45, 46], might be readily testable in terms of the apparent charge fraction, f. Furthermore, a detailed atomic level understanding of the differing roles of monovalent and divalent ions may be gleaned from MD simulations, once discrepancies with experiment are resolved. MD simulations will also be able elucidate the presence of counterion-induced deformations in the chain, or 'wrinkles', that have been observed in coarse-grained MD with explicit ions [47].

Finally, and most significantly, we plan to extend these approaches to study chains of mixed sequence, to gain insight into the sequence preferences found in important biological molecules such as mRNA or riboswitches. Sequence and sugar (ribose vs. deoxyribose) effects are expected to change base stacking interactions, which adds another level of complexity, but in principle can also be modeled using virtual bonds with measured B_2. Tight collaboration with MD will be essential to explore the subtle details that lead to highly specialized biological function.

Acknowledgements

We thank Pollack group members and CHESS scientists Ken Finkelstein and Arthur Woll, and staff, for experimental assistance. This work was supported by the NIH (R01-GM085062). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0936384, and the Mac-

CHESS resource is supported by the National Institute of General Medical Services (NIGMS) award GM-103485.

2.5 Materials and Methods

2.5.1 Preparation of dT_{30} samples for x-ray scattering experiments

$DT_{30}$was synthesized and HPLC purified by Integrated DNA Technologies, Inc. For SAXS studies of the conformation and electrostatics in mixed ion environments, dT_{30} was buffer exchanged with solutions containing 1 mM Na-MOPS, pH 7.0, 20 mM NaCl and 0, 1, 2, 5, 10, or 20 mM MgCl2 using spin concentrators (Amicon Ultra-0.5, 3kDa cutoff, EMD Millipore). For each ionic condition, a DNA concentration series was prepared by dilution with the matching buffer. For measurements of the ion atmosphere, dT_{30} was buffer exchanged with 1 mM Na MOPS pH 7.0 and 100 mM NaCl or 100 mM RbCl. The DNA concentration was determined by UV absorption at room temperature (Cary 50, Varian Inc.) assuming an extinction coefficient of $\varepsilon_{260\ \mathrm{nm}} = 2.436 \times 10^5 \mathrm{M}^{-1}\mathrm{cm}^{-1}$.

2.5.2 SAXS data collection and analysis

SAXS data were collected at CHESS beamline G1 using an x-ray energy of 10.53 keV. Samples and matching buffers were loaded sequentially in a 2 mm diameter, 10 μm thick quartz capillary. During each exposure, scattering patterns were collected on a photon-counting area detector (Pilatus 100K, Dectris) at a distance of 1.689 m from the sample and normalized by the beamstop PIN diode current. Oscillation of the sample within the capillary was used to reduce radiation damage [48]. Data processing was performed in MATLAB. X-ray images were az-

imuthally averaged, and repeated exposures of the same sample were compared to verify that no damage had occurred. The uncertainty of $I(q_i)$ at each bin of N_i pixels was estimated from the standard deviation $\sigma_i$as $\sigma_i N_i^{-1/2}$. This uncertainty was propagated through buffer subtraction and averaging of repeated exposures.

SAXS profiles at varying DNA concentration c were matched over the range $0.1 < q < 0.26\text{Å}^{-1}$, and the form factor $P(q)$ was obtained by linear extrapolation to $c = 0$ of $I(q,c)$ at each value of q. In order to determine the second virial coefficient, B_2, these set of curves were fit using a model for the interparticle interference of dilute solutions,

$$I(q,c) = P(q) - 2P(0)B_2A_0cP_{IPI}(q) \tag{2.13}$$

where $P_{IPI}(q)$ is the q-dependent part of the interparticle interference term with $P_{IPI}(0) = 1$, and A_0 is defined so that c is in M and B_2 is in Å^3. In order to find $P(0)$, the low-q portion of $P(q)$ was fit with the Debye function for a random coil $P_D(q)$ [49]. The interference terms $|I(q,c)–P(q)|$ were found empirically to have a Gaussian shape for $q < 0.035\text{Å}^{-1}$. The Gaussian shape is also observed in theoretical calculations of the interference function (see Fig 2.11). Therefore, we chose $P_{IPI}(q) = \exp(-q^2d^2)$, where the length scale d is independent DNA concentration, but may depend on salt. For each salt condition, we obtained B_2 (and d) using a simultaneous nonlinear least squares fit of the interference terms to $2P(0)B_2A_0cP_{IPI}(q,d)$ for $q < 0.035\text{Å}^{-1}$. In order to double-check the validity of the above assumptions, the low-q profiles were reconstructed using $I(q,c) = P_D(q) - 2P(0)B_2A_0c\exp(-q^2d^2)$ and compared directly with the experimental data, as in the inset of Fig. 2.2.

2.5.3 Heavy ion and ASAXS measurement of excess ions around dT_{30}

For the heavy ion replacement method, solutions of 0.1 mM dT_{30} in NaCl and RbCl were prepared as described above, and placed on an absolute scale using water as a calibrant [50]. The largest experimental uncertainty for this technique was from the measurement of the DNA concentrations. Therefore, ΔN^e_{tot} was calculated separately for repeated concentration measurements, and the uncertainty in N_{ions} was estimated using linear regression.

ASAXS profiles from 0.3 mM dT_{30} with Rb were acquired at beamline C1 at CHESS, as described in [29]. Two x-ray energies were chosen below the Rb absorption edge, $E_{lo} = 15.093$ and $E_{on} = 15.191$ keV, and the x-ray fluorescence spectrum from a dilute RbCl solution was measured and used to calculate f' at E_{lo} and E_{on} [51]. The number of Rb ions was obtained using $N_{Rb} = (I_{lo}(0)^{1/2} – I_{on}(0)^{1/2}) S(0)^{-1/2} \Delta f'^{-1}$, where $S(0) = 0.95$ corrects for the interparticle interference effect [29]. The extrapolations of $I_{lo}(q)$ and $I_{on}(q)$ to $q = 0$ shown in Fig. 2.5 were performed using the Bayesian indirect Fourier transform, as described above.

2.5.4 Single molecule FRET

For smFRET measurements of the end to end distance, dT_{30} with 3' Cy3 and 5' Cy5 labels was synthesized and HPLC purified by Integrated DNA Technologies, Inc.. DNA was prepared in solutions containing 20 mM TRIS buffer, pH 8, 20 mM NaCl and 0, 0.5, 1, 5, 10, or 20 mM $MgCl_2$.

The experimental methods were similar to those used in a previous smFRET study involving poly(dT) [8]. Briefly, smFRET measurements of the freely diffusing DNA were performed on a custom confocal microscope using an Olympus UAPO 40x objective. Donor and acceptor fluorescence were separated by a 660LP dichroic, then additionally filtered by a 570/40 band-pass

(donor channel) and 660 long pass (acceptor channel). Each fluorescence signal was collected through a 30 μm pinhole and detected by a single-photon counting PMT (Hamamatsu). The intensity was sampled at 60MHz using a Flex03LQ-01 correlator card (correlator.com) in the photon counting mode.

Data analysis was also similar previous work [8], one exception being that a threshold was only placed on the sum of donor and acceptor channels, rather than on each channel individually. Histograms of events passing the threshold criterion were fit to two Gaussians: the first low-FRET peak representing molecules with an inactive acceptor, and the second representing the signal of interest. The peak FRET efficiency E_{FRET} was converted to the FRET-averaged end-to-end distance $\langle R \rangle_{\mathrm{FRET}} = R_0(1/E_{\mathrm{FRET}} - 1)^{1/6}$ using the experimentally-derived Förster radius for the dye pair conjugated to poly(dT): $R_0 = 56.4$ Å [8].

2.5.5 Calculation of scattering profiles from the virtual bond representation of poly(dT)

For a chain with contour length L and a uniform density along the contour, the form factor is

$$P_{\mathrm{chain}}(q) = \int_0^1 \int_0^1 ds\, ds' \, \frac{\sin(q\,|\vec{r}(s) - \vec{r}(s')|)}{q\,|\vec{r}(s) - \vec{r}(s')|} \tag{2.14}$$

where s is the distance along the chain divided by the contour length. To efficiently calculate $P_{\mathrm{chain}}(q)$, a histogram of distances between points was generated and $P_{\mathrm{chain}}(q)$ was found using Equation 2.10. For comparison between chain form factors and CRYSOL calculations (Fig. 2.7), s was discretized with $L\Delta s \approx 0.1$Å, and the histogram was generated for all pairs of points. When computing form factors for a distribution of chain conformations (Fig. 2.10), the histogram was generated from randomly sampled chains and values of s on the interval [0,1].

2.5.6 Calculation of the interparticle interference function

Calculation of the interparticle interference function uses a Monte Carlo method [41], and begins as described above for the calculation of B_2 by generating pairs of chains drawn from$\Omega_{\kappa,f}$ and center of mass displacement vectors with uniform probability density in a volume $V_{\max}$. For each pair of chains labeled by the index m, an inter-particle distance histogram $p_i^{(m)}$ with bins centered at r_i is calculated for distance vectors that span the two chains, normalized so that $\sum_i p_i^{(m)} = 1$. The interaction energy between chains W_m was calculated as described above. After M iterations, the interparticle interference function was computed using

$$I_{\mathrm{IPI}}(q) = A_0 c \frac{V_{max}}{M} \sum_{i,m} p_i^{(m)} (1 - \exp(-W_m)) \frac{\sin(q\, r_i)}{q\, r_i} \tag{2.15}$$

2.5.7 Monte Carlo integration of the ion atmosphere

At each step in the Monte Carlo integration, a chain was chosen from the distribution $\Omega_{\kappa,f}$ and a point was chosen uniformly within the integration volume (a sphere of radius $r_{\max} = 160$ Å where the origin is the chain's center of mass). The point was classified as being within one of three phases according to the electrostatic potential $\Phi(\vec{r})$ defined in equation 2.12, and the distance between the point at and the nearest chain atom: $d_{\min}$,

$$\text{phase at } \vec{r} = \begin{cases} \text{chain}, & d_{\min} < a \\ \text{condensed ion}, & d_{\min} \geq a \text{ and } \Phi(\vec{r}) \geq \Phi_{\text{cutoff}} \\ \text{diffuse ion}, & d_{\min} \geq a \text{ and } \Phi(\vec{r}) < \Phi_{\text{cutoff}} \end{cases} \tag{2.16}$$

The chain and condensed ion phases were assigned densities of 1 in order to compute their

volumes, and the density for the diffuse ion phase was calculated from the potential through Equation 2.11. A radial density histogram for each phase was incremented accordingly. After 10^6 iterations, the number of diffuse ions within a radius r of the chain center of mass was determined from a cumulative sum of the density histogram. Similarly, the volumes of the chain and condensed ion phases within r, $V(r)$, were given by the sum of the respective density histograms. The number within r was calculated assuming uniform density of particles within these phases: $N(r) = N_{\text{tot}} V(r)/V(r_{\text{max}})$.

2.5.8 Explicit solvent MD simulations of $\mathbf{dT}_{10}$

150ns of Molecular dynamics simulations at 300K was performed with MOIL suite of programs [52, 53] with explicit account of water, ions and DNA. To model the nucleic acid and its environment we used TIP3P [54] for water together with recent AMBER parameters for monovalent ions [55] and DNA [56, 57]. The simulated system contained dT_{10}, 5062 water molecules, and 10 Na^+ and 1 Cl^- ions. The center of mass of the DNA was constrained by harmonic springs to the center of the simulation box. The size of the simulation box was (55.75 Å)^3, and periodic boundary conditions were applied in all directions. The long-range component of the electrostatics interaction was calculated by particle mesh Ewald method [58] with a grid spacing of $32 \times 32 \times 32$ Å. The cutoff for the real space part of the electrostatics and van der Waals interactions was set to 8.5 Å with a non-bonded list update for every 8 steps. We used matrix version [59, 60] of the SHAKE algorithm [61] to constrain the water bond lengths and angles; the bond lengths of the DNA fixed by SHAKE that allowed us to use 1.5 fs time step. We also used RESPA [62] for dual time stepping. Here the reciprocal-space of the Ewald sum was calculated every four steps, while the rest of the forces were evaluated every step. We reported atom positions for every 3 ps for further analysis.

To compute the excess ions around the flexible chain (dT_{10}) we cut a slice of width d from the outer face of the cubic box in all directions. These regions are combined and used to compute the asymptotic concentration of Na^+ ions (c_{bulk}). Once we compute the c_{bulk} we find the excess ions simply as$N_{excess} = N_{Na+} - V_{box}c_{bulk}$ where N_{Na^+} is the total number of Na^+ ions (10 here) and V_{box} is the total volume of the simulation box. The width of the cut d is an *ad hoc* parameter. We used $d = 3$, 4, 5 and 6 Å for comparison. The results for cbulk with alternate cutoffs differ by no more than 10% suggesting convergence in estimating the bulk concentration. As a final result we reported the average of these results.

References

[1] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules", Proc. Natl. Acad. Sci. U.S.A. **94**, 6185–6190 (1997).

[2] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins", Nature **450**, 964–972 (2007).

[3] C. D. Eichhorn, J. Feng, K. C. Suddala, N. G. Walter, C. L. Brooks, and H. M. Al-Hashimi, "Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch", Nucleic Acids Res. **40**, 1345–1355 (2012).

[4] M. J. Campolongo, S. J. Tan, D.-M. Smilgies, M. Zhao, Y. Chen, I. Xhangolli, W. Cheng, and D. Luo, "Crystalline Gibbs monolayers of DNA-capped nanoparticles at the air-liquid interface", ACS Nano **5**, 7978–7985 (2011).

[5] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy", Biophys. J. **86**, 2530–2537 (2004).

[6] T. A. Laurence, X. Kong, M. Jäger, and S. Weiss, "Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins", Proc. Natl. Acad. Sci. U.S.A. **102**, 17348–17353 (2005).

[7] S. Doose, H. Barsch, and M. Sauer, "Polymer properties of polythymine as revealed by translational diffusion", Biophys. J. **93**, 1224–1234 (2007).

[8] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA", Proc. Natl. Acad. Sci. U.S.A. **109**, 799–804 (2012).

[9] M. N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, "Stretching single stranded DNA, a model polyelectrolyte", Phys. Rev. Lett. **89**, 248102+ (2002).

[10] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck, "Nonlinear low-force elasticity of single-stranded DNA molecules", Phys. Rev. Lett. **102**, 068301 (2009).

[11] Y. Seol, G. M. Skinner, and K. Visscher, "Elastic properties of a single-stranded charged homopolymeric ribonucleotide", Phys. Rev. Lett. **93**, 118102+ (2004).

[12] A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, "Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle X-ray scattering", Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. **86**, 021901+ (2012).

[13] B. Y. Ha and D. Thirumalai, "Electrostatic persistence length of a polyelectrolyte chain", Macromolecules **28**, 577–581 (1995).

[14] T. Odijk, "Polyelectrolytes near the rod limit", J. Polym. Sci. Polym. Phys. Ed. **15**, 477–483 (1977).

[15] J. Skolnick and M. Fixman, "Electrostatic persistence length of a wormlike polyelectrolyte", Macromolecules **10**, 944–948 (1977).

[16] J. L. Barrat and J. F. Joanny, "Persistence length of polyelectrolyte chains", Europhys. Lett. **24**, 333+ (2007).

[17] P. J. Flory, *Principles of polymer chemistry* (Cornell University Press, 1953).

[18] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, and B. Schuler, "Charge interactions can dominate the dimensions of intrinsically disordered proteins", Proc. Natl. Acad. Sci. U.S.A. **107**, 14609–14614 (2010).

[19] A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu, "Net charge per residue modulates conformational ensembles of intrinsically disordered proteins", Proc. Natl. Acad. Sci. U.S.A. **107**, 8183–8188 (2010).

[20] S. Cao and S.-J. Chen, "Predicting RNA folding thermodynamics with a reduced chain representation model", RNA **11**, 1884–1897 (2005).

[21] W. K. Olson and P. J. Flory, "Spatial configurations of polynucleotide chains. I. Steric interactions in polyribonucleotides: a virtual bond model", Biopolymers **11**, 1–23 (1972).

[22] X. Qiu, L. W. Kwok, H. Y. Park, J. S. Lamb, K. Andresen, and L. Pollack, "Measuring Inter-DNA potentials in solution", Phys. Rev. Lett. **96**, 138101+ (2006).

[23] S. Kirmizialtin and R. Elber, "Computational exploration of mobile ion distributions around RNA duplex", J. Phys. Chem. B **114**, 8207–8220 (2010).

[24] S. Kirmizialtin, S. A. Pabit, S. P. Meisburger, L. Pollack, and R. Elber, "RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations", Biophys. J. **102**, 819–828 (2012).

[25] X. Qiu, K. Andresen, L. W. Kwok, J. S. Lamb, H. Y. Park, and L. Pollack, "Inter-DNA attraction mediated by divalent counterions", Phys. Rev. Lett. **99**, 038104+ (2007).

[26] G. Porod, "General theory", in *Small angle x-ray scattering*, edited by O. Glatter and O. Kratky (Academic Press, Inc., New York, 1982), pp. 18–50.

[27] S. Hansen, "Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering", J. Appl. Crystallogr. **33**, 1415–1421 (2000).

[28] R. Das, T. Mills, L. Kwok, G. Maskel, I. Millett, S. Doniach, K. Finkelstein, D. Herschlag, and L. Pollack, "Counterion distribution around DNA probed by solution X-ray scattering", Phys. Rev. Lett. **90**, 188103 (2003).

[29] S. A. Pabit, S. P. Meisburger, L. Li, J. M. Blose, C. D. Jones, and L. Pollack, "Counting ions around DNA with anomalous small-angle X-ray scattering", J. Am. Chem. Soc. **132**, 16334–16336 (2010).

[30] Y. Marcus, *Ion solvation* (Wiley, 1985).

[31] G. Lamm and G. R. Pack, "Counterion condensation and shape within Poisson-Boltzmann theory", Biopolymers **93**, 619–639 (2010).

[32] Y. Zhang, H. Zhou, and Z. C. Ou-Yang, "Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions.", Biophys. J. **81**, 1133–1143 (2001).

[33] E. J. W. Verwey and J. T. G. Overbeek, *Theory of the stability of lyophobic colloids* (Elsevier, 1948).

[34] T. Yadav, B. Carrasco, A. R. Myers, N. P. George, J. L. Keck, and J. C. Alonso, "Genetic recombination in Bacillus subtilis: a division of labor between two single-strand DNA-binding proteins", Nucleic Acids Res. **40**, 5546–5559 (2012).

[35] K.-W. Chan, Y.-J. Lee, C.-H. Wang, H. Huang, and Y.-J. Sun, "Single-stranded DNA-binding protein complex from Helicobacter pylori suggests an ssDNA-binding surface", J. Mol. Biol. **388**, 508–519 (2009).

[36] J. Fan and N. P. Pavletich, "Structure and conformational change of a replication protein A heterotrimer bound to ssDNA", Genes Dev. **26**, 2337–2347 (2012).

[37] S. Kamtekar, A. J. Berman, J. Wang, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, and T. A. Steitz, "Insights into strand displacement and processivity from the crystal structure

of the protein-primed DNA polymerase of bacteriophage Φ29", Mol. Cell **16**, 609–618 (2004).

[38] D. Svergun, C. Barberato, and M. H. J. Koch, "CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates", J. Appl. Crystallogr. **28**, 768–773 (1995).

[39] P.-G. De Gennes, *Scaling concepts in polymer physics* (Cornell university press, 1979).

[40] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines", J. Chem. Phys. **21**, 1087–1092 (1953).

[41] S. R. Hubbard and S. Doniach, "A Monte Carlo calculation of the interparticle interference in small-angle X-ray scattering", J. Appl. Crystallogr. **21**, 953–959 (1988).

[42] Z. Alexandrowicz and A. Katchalsky, "Colligative properties of polyelectrolyte solutions in excess of salt", J. Polym. Sci. A **1**, 3231–3260 (1963).

[43] D. E. Makarov and K. W. Plaxco, "Measuring distances within unfolded biopolymers using fluorescence resonance energy transfer: the effect of polymer chain dynamics on the observed fluorescence resonance energy transfer efficiency", J. Chem. Phys. **131**, 085105+ (2009).

[44] Y. Bai, M. Greenfeld, K. J. Travers, V. B. Chu, J. Lipfert, S. Doniach, and D. Herschlag, "Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids", J. Am. Chem. Soc. **129**, 14981–14988 (2007).

[45] A. Kundagrami and M. Muthukumar, "Theory of competitive counterion adsorption on flexible polyelectrolytes: divalent salts", J. Chem. Phys. **128**, 244901+ (2008).

[46] Z.-J. J. Tan and S.-J. J. Chen, "RNA helix stability in mixed Na^+/Mg^{2+} solution.", Biophys. J. **92**, 3615–3632 (2007).

[47] M. J. Stevens, D. B. McIntosh, and O. A. Saleh, "Simulations of stretching a strong, flexible polyelectrolyte", Macromolecules **45**, 5757–5765 (2012).

[48] S. S. Nielsen, M. Møller, and R. E. Gillilan, "High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell", J. Appl. Crystallogr. **45**, 213–223 (2012).

[49] P. Debye, "Molecular-weight determination by light scattering.", J. Phys. Chem. **51**, 18–32 (1947).

[50] D. Orthaber, A. Bergmann, and O. Glatter, "SAXS experiments on absolute scale with Kratky systems using water as a secondary standard", J. Appl. Crystallogr. **33**, 218–225 (2000).

[51] G. Evans and R. Pettifer, "CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra", J. Appl. Crystallogr. **34**, 82–86 (2001).

[52] R. Elber, A. Roitberg, C. Simmerling, R. Goldstein, H. Li, G. Verkhivker, C. Keasar, J. Zhang, and A. Ulitsky, "MOIL: a program for simulations of macromolecules", Comput. Phys. Commun. **91**, 159–189 (1995).

[53] A. P. Ruymgaart, A. E. Cardenas, and R. Elber, "MOIL-opt: Energy-Conserving molecular dynamics on a GPU/CPU system", J. Chem. Theory Comput. **7**, 3072–3082 (2011).

[54] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", J. Chem. Phys. **79**, 926–935 (1983).

[55] I. S. Joung and T. E. Cheatham, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations", J. Phys. Chem. B **112**, 9020–9041 (2008).

[56] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", J. Am. Chem. Soc. **117**, 5179–5197 (1995).

[57] M. Zgarbová, M. Otyepka, J. Sponer, A. Mládek, P. Banáš, T. E. Cheatham, and P. Jurečka, "Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles.", J. Chem. Theory Comput. **7**, 2886–2902 (2011).

[58] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method", J. Chem. Phys. **103**, 8577–8593 (1995).

[59] R. Elber, A. P. Ruymgaart, and B. Hess, "SHAKE parallelization", Eur. Phys. J. Spec. Top. **200**, 211–223 (2011).

[60] Y. Weinbach and R. Elber, "Revisiting and parallelizing SHAKE", J. Comput. Phys. **209**, 193–206 (2005).

[61] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes", J. Comput. Phys. **23**, 327–341 (1977).

[62] M. Tuckerman, B. J. Berne, and G. J. Martyna, "Reversible multiple time scale molecular dynamics", J. Chem. Phys. **97**, 1990–2001 (1992).

CHAPTER 3

X-RAY SCATTERING DATA FROM POLYDEOXYTHYMIDINE FOR MOLECULAR DYNAMICS FORCE FIELD VALIDATION

Steve P. Meisburger[1], Serdal Kirmizialtin[2,3], Ron Elber[3], and Lois Pollack[1]

1. Applied and Engineering Physics, Cornell University, Ithaca, New York
2. Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratories, Los Alamos, New Mexico
3. Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas, USA.

Abstract

Molecular dynamics (MD) is a powerful tool for understanding the complex network of interactions that constitute biomolecular function. The discovery of new roles for RNA and DNA in living organisms has created a need for predictive, accurate MD simulations of nucleic acids. Nucleic acids pose distinct challenges for simulation because of their strong interactions with salt ions, coupling of backbone torsion angles, and ubiquitous pi-stacking interactions. Within the framework of additive force fields such as AMBER, parameters developed for proteins have been re-tuned to reproduce some essential features of nucleic acids, such as the well-studied B-form of DNA. However, success in demanding applications such as *de-novo* folding of RNA tertiary structure has been rare. Recently, Chen & Garcia [1] showed that productive *de-novo* folding of RNA in-silico was hampered by off-pathway collapsed states, due in part to an overly strong base-stacking interaction in AMBER-99. Thus, for accurate prediction of nucleic acid

structure and dynamics, it is important to benchmark the performance of MD force fields using structural data from the *unfolded* state of nucleic acids.

In order to identify parameter sets that provide balanced representations of the unfolded states, we propose that simulations of short sequences lacking secondary structure should be directly compared with high quality solution X-ray scattering data from identical systems. As proof of principal, we collect small and wide-angle x-ray scattering data from a model unfolded ssDNA homopolymer dT_{10} in neutral pH buffer with 100 mM monovalent salt, and over a range of temperatures from 4 to 40C. The same system is simulated using explicit solvent MD for three different nucleic acid force fields. Strikingly, these force fields universally converge to compact ensembles that do not agree with experimental data at any temperature. In contrast, a coarse-grained polyelectrolyte model including only repulsive interactions reproduces available SAXS data for poly(dT) molecules of 10-40 nt (dT_{10}-dT_{40}). This suggests that the collapsed ssDNA ensembles commonly seen in MD simulations are an artifact of force-field imbalance. Re-tuning efforts currently underway will be benchmarked using the proposed methodology.

3.1 Introduction

Simulation of atomic motions using Newtonian mechanics is the workhorse technique for biophysical modeling of proteins. With the ongoing discovery of new roles for nucleic acids in biology, there is a pressing need for accurate and predictive simulation of nucleic acid folding and dynamics for applications related to human health. RNA molecules have enzymatic [2] and regulatory functions [3–5] that are essential for normal function of mammalian cells. Additionally, RNA-based machines unique to pathogenic bacteria [6, 7] and viruses [8, 9] are important drug targets. Biotechnology applications such as nanopore sequencing [10] also rely on accurate structure prediction for single-stranded DNA [11]. However, well-established force

fields originally optimized for protein simulation often fail to reproduce important structural properties of nucleic acids [1, 12].

Here, we consider the challenge of simulating unfolded nucleic acids, an important intermediate in structural transitions. De-novo folding simulations have shown that accurate simulation of the unfolded state is essential for productive folding: if the force-field is not properly tuned, collapsed intermediates trap the simulation in non-productive trajectories [1]. Thus, we propose that the dimensions of intrinsically unstructured nucleic acids should provide a valuable benchmark for molecular dynamics.

Simulating unfolded nucleic acids is currently a challenge. Unfolded nucleic acids lack the strong interactions that stabilize the native fold, and therefore the global conformation is highly sensitive to torsion angle potentials, electrostatic forces, and molecular hydration. There are many indications that MD simulations do not accurately capture these forces. Torsion angle parameters are "the Achilles heel of force field computations" [13] and are poorly constrained by chemical data. Interactions between positive ions and negatively-charged nucleic acids are important determinants of structure [14] and couple to solvent degrees of freedom [15]. Thus, the ion and water models must also be balanced and accurate. Commonly-used water models may not be sufficiently accurate to predict nucleic acid dynamics: for example, TIP3P has poor accuracy with respect to bulk properties such as viscosity [16]. Finally, simulating unfolded random coils is computationally challenging because large simulation boxes and long simulation times are required to sample the molecular and solvent degrees of freedom completely.

On the experimental side, there is a lack of reliable structural data for unfolded nucleic acids that would constitute a rigorous test of MD. High resolution structural techniques such as 2D-NMR and x-ray crystallography suffer significant loss of information when disorder is present. In contrast, small angle scattering does not suffer a loss of signal due to disorder, beyond the added ambiguity of interpreting the ensemble-averaged intensity. Much of the previous work

on x-ray scattering from ssDNA has relied on simple polymer models for interpretation [17–20]. The wormlike chain model [21] is commonly invoked because the number of parameters is small and the length-dependence of the radius of gyration can be predicted analytically [18, 19]. The wormlike chain model has been modified to include an electrostatic, salt-dependent flexibility [22–25]. However, the model neglects excluded volume interactions that predominate in highly flexible chains like ssDNA at low ionic strengths [26]. Thus, even if parameters may be found to fit some data sets at low resolution (e.g. Refs. [18, 19]), the model offers limited insight into the physical origin of electrostatic effects.

More recently, we used coarse-grained simulation of a ssDNA-like polymer consisting of rigid virtual bonds (two bonds per nucleotide) interacting through a screened Yukawa potential to interpret the scattering data [20]. In this case, the interaction potential between molecules was measured by SAXS to understand the role of ions in charge screening. Once charge screening was properly accounted for in the polymer model, the overall extent of chain expansion agreed well with the available data. Agreement with the full scattering profiles was also very good, but not exact. The rearrangement of the backbone that may accompany ion binding was not captured in the model, nor were the bases explicitly modeled. Important questions remain concerning specific ion effects that will benefit from accurate atomistic simulation.

Here, we lay the groundwork for a detailed comparison between all-atom MD and SAXS for the unfolded states of RNA and DNA. First, we provide reliable SAXS data for molecules that are small enough to allow computationally feasible all-atom MD simulations (dT_{10} in 100 mM NaCl). Scattering curves were collected over a large range in q to capture high-resolution features. In addition, data were acquired at multiple temperatures, facilitating comparison with replica-exchange MD simulations. Second, we clarify the polymer scaling properties of poly(dT), so that features observed in MD simulations of short chains may be placed in the context of previous studies on longer chains. Finally, as proof-of-principle, MD simulation results

for dT10 are presented. Three force field parameter sets are compared, including OPLSAA and two variants of AMBER-99. It is obvious from the computed SAXS profiles that these simulations do not agree with the experiment: MD simulations show various degrees of structural collapse that are inconsistent with the experimentally measured dimensions of the molecule in solution. We find no evidence that dT10 forms such collapsed states under these ionic conditions, even at low temperature. The existence of a collapsed state, ubiquitously seen in MD simulations of ssDNA (e.g. [27, 28]), is also inconsistent with the experimental observation of random-coil-like scaling behavior of the size with the number of monomers. In contrast, very good agreement is obtained using a coarse-grained polymer model that includes only repulsive interactions. We conclude that SAXS data for poly(dT) will be a useful benchmark for new force field tuning efforts currently underway.

3.2 Results and Discussion

3.2.1 Poly(dT) in 100mM NaCl has dimensions of a random, swollen coil

DT_{10} molecules lacking the 5' terminal phosphate were chemically synthesized by Integrated DNA Technologies (Coralville, IA). Samples were buffer-exchanged with 100 mM NaCl, 1 mM Na+MOPS, pH 7.0, as in previous studies [18, 20]. SAXS curves for dT_{10} were measured in dilute solution (0.72 mg mL^{-1}) at MacCHESS beamline F2 [29], in a SAXS/WAXS configuration using two Pilatus 100K detectors to achieve an extended q-range up to 0.8 $Å^{-1}$ ($q = 4\pi \sin(2\theta/2)/\lambda$ where 2θ is the scattering angle and λ is the wavelength). The scattering signal for dT_{10} was weak compared with longer homopolymers because of its low molecular weight, however a reasonable signal-to-noise ratio was achieved by taking long exposures and continuously oscillating the sample in the cell using a syringe pump to distribute the dose.

The scattering from dT_{10} was compared visually with previous data for dT_{30} [20] and dT_{40} [18] acquired under identical solution conditions (but at different x-ray beamlines). Agreement of the scattering data at high angle for varying length is characteristic of linear polymers. For all three scattering curves, the radius of gyration (Rg) was determined by the indirect Fourier transform (IFT) method [30] with Bayesian estimation of all free parameters [31]. We find $R_G = 32.7 \pm 0.8$ Å for dT_{40}, 27.6 ± 0.2 Å for dT_{30}, and 12.8 ± 0.2 Å for dT_{10}.

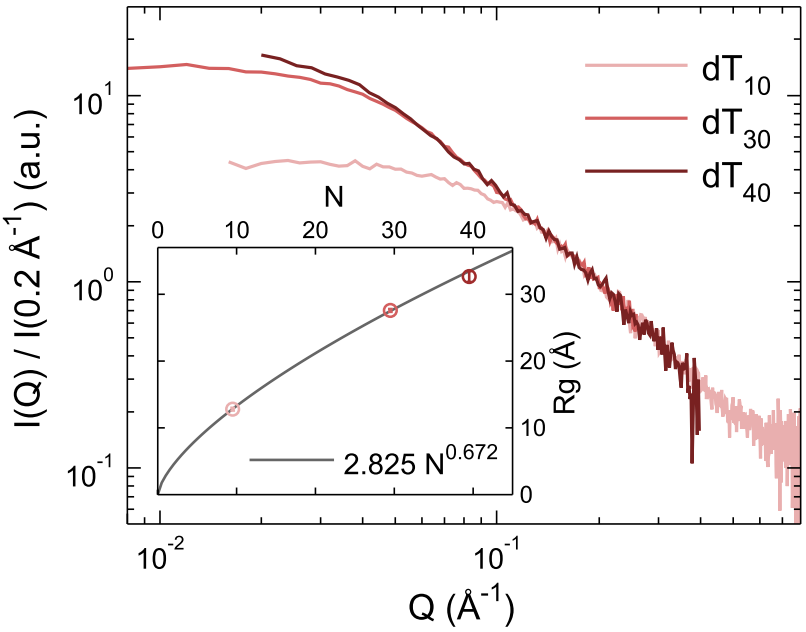

Figure 3.1: SAXS data for poly(dT) at room temperature in buffer containing 100mM NaCl, 1 mM Na-MOPS, pH 7.0. Data for dT10 (present work), dT30 [20], and dT40 [18] were normalized by the value at $q = 0.2\text{Å}^{-1}$. Inset: Radius of Gyration (Rg) vs. number of monomers (N) was fit to a power law (solid line) with numerical values given in the legend. For error estimates on the power law parameters, see the text.

In polymer theory, the predicted dependence of Rg on length N follows a power law

$$R_G = aN^{\nu} \tag{3.1}$$

where $\nu = 0.5$ characterizes a non-interacting Gaussian chain, $\nu > 0.5$ signifies a self-avoiding polymer, and $\nu < 0.5$ characterizes collapsed chains [32]. The upper limit $\nu = 1$ describes an extended rod-like state, and the lower limit $\nu = 0.33$ describes a fully-collapsed state. Because

the synthetic ssDNA constructs lack electron-rich 5' phosphate groups, we assumed $N = 9.5$ for dT_{10}, $N = 29.5$ for dT_{30}, and $N = 39.5$ for dT_{40}. The fit of R_G vs. N to Equation 3.1 is shown in the inset of Figure 3.1. The best-fit values are $a = 2.82 \pm 0.14$ and $\nu = 0.672 \pm 0.015$. From the above polymer scaling arguments, $\nu > 0.5$ suggests that poly(dT) is a self-avoiding random coil.

A similar scaling analysis has been reported for poly(dT) in comparable solution conditions in Ref. [19], and $\nu > 0.5$ was also observed. The precise values of a and ν appear to be at odds with our findings. However, the SAXS data in Ref. [19] were not always acquired over a sufficient q-range to determine R_G reliably, except by fitting to the Debye formula which presupposes Gaussian coil statistics. If we use the Debye formula rather than IFT to determine R_G, we find values of a and ν consistent with previous work. Note, however, that IFT is a relatively unbiased method for finding R_G, and should be preferred as long as the minimum q value is below$\sim \pi/d_{max}$ (which is always the case in the present analysis).

Finally, analysis of the shape of the scattering curve via a Kratky plot (q vs $q^2I(q)$) can provide qualitative information about the underlying scaling laws [33]. In Figure 3.2, the high-q scaling for dT_{10} is intermediate between that of a compact object (sphere), and a fully extended one (thin rod).

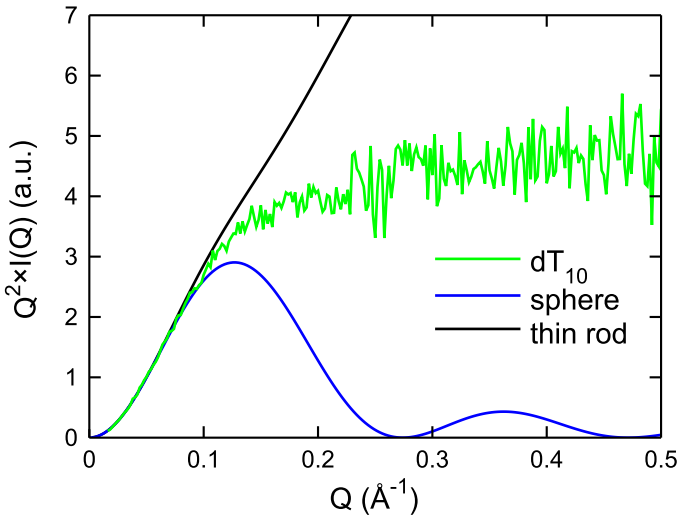

Figure 3.2: SAXS data for dT_{10} (20 C) shown in the Kratky presentation, with axes of Q vs $Q^2 \times I(q)$ to emphasize power law scaling. Drawn on the same axes are the theoretical scattering profiles for a solid sphere (blue) and a thin rod (black) having the same radius of gyration and forward scattering as dT_{10} (green).

Given that the power law scaling is suggestive of a self-avoiding polymer, excluded volume effects should be taken into account when modeling the conformations of poly(dT). We recently developed a polyelectrolyte model that represents poly(dT) by a virtual-bond freely-rotating chain, and includes salt-dependent electrostatic interactions determined from measurements of the second virial coefficient for semi-dilute solutions (which enter into the model via a parameter f giving the re-normalization of the surface charge due to condensed counterions) [20]. The parameter f was set according to measurements of dT_{30} in 20 mM NaCl, a lower salt concentration than used here for dT_{10}. In order to predict scattering curves for the present data set, we assume that the same value of f applies for 100 mM NaCl, and for varying lengths of the DNA. Otherwise, the model is adopted without modification (i.e. with no free parameters except for the overall intensity). Briefly, self-avoiding chains were generated with numbers of phosphates (bases) equal to 9 (10), 29 (30), and 39 (40) with geometric parameters from Ref. [20]: virtual bond length $l_0 = 3.69\,\text{Å}$, valence angle $\theta = 57.75^\circ$, and chain diameter $d = 5.6\,\text{Å}$. Electrostatic repulsion was calculated assuming an ionic strength $I = 0.1\,\text{M}$ and charge renormalization pa-

rameter $f = 0.6$. The bare chain scattering factor was multiplied by an effective cross-section, parametrized using CRYSOL calculations from crystal structures of poly(dT) bound to protein, as $I_{CS}(q) = (1-c) \times (2J_1(qR)/(qR))^2 + c$ where $R = 6.812$Å, $c = 0.2838$, and $J_1(qR)$ is a Bessel function of the first kind [20]. This effective cross-sectional scattering factor is expected to be a good approximation only out to $q \sim 0.25\,\text{Å}^{-1}$: i.e. beyond this value, it is important to explicitly model the structure of the nucleotides.

The scattering curve predictions from the model are compared with the data for dT_{10}, dT_{30}, and dT_{40} in Figure 3.3. The radii of gyration predicted by model are 12.4, 27.6, and 34.1 Å, all within a few percent of the measured values. Furthermore, the shapes of the scattering curves show excellent agreement at large scattering angles, suggesting that the polymer statistics are modeled correctly in all three cases. Importantly, the model includes no attractive interactions: only the hard-core excluded volume and repulsive screened coulomb potential are included. Collapsed conformations are not significantly populated in the simulated conformational ensemble. Thus, the solution structure of poly(dT) is consistent with a self-avoiding random coil.

3.2.2 The random coil state of poly(dT) persists over a wide temperature range

The temperature-controlled sample cell available at MacCHESS beamline F2 allows data collection with a temperature range of 4-40 C. The in-vacuum sample cell prevents water condensation on the cell window, a problem for in-air cells when cooled to low temperatures. Scattering curves for dT_{10} were measured at 4, 20, and 40 C, and subtracted from buffer "blanks" taken at the same temperatures. From the data, shown in Figure 3.4 in the Guinier representation, we observe no significant variation with temperature. Radii of gyration determined by a direct linear fit in the Guinier region ($q < 1.3R_G^{-1}$) are the same within error (Figure 3.4, inset). Since

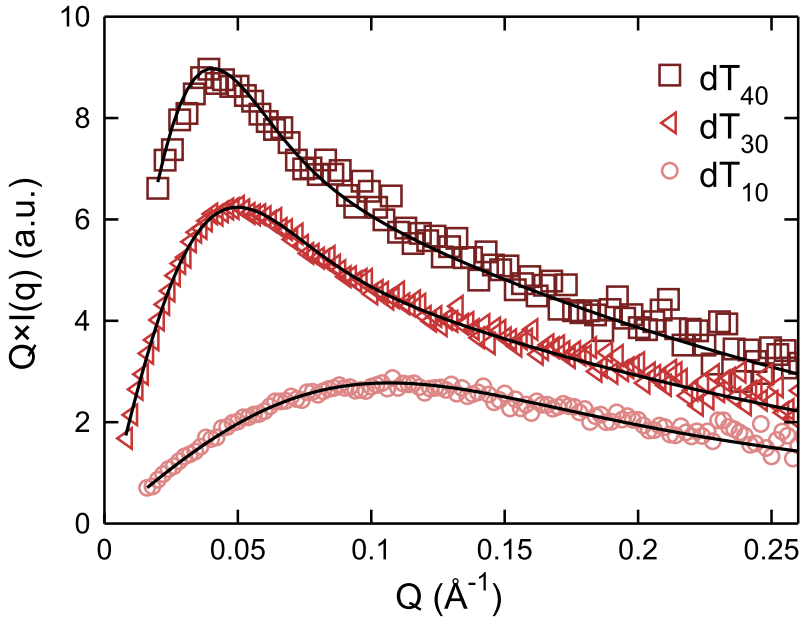

Figure 3.3: Form factor predictions from a coarse-grained polymer model for poly(dT) [20] are shown with the SAXS data from Figure 3.1. The only adjusted parameter in this model is the (arbitrary) overall intensity. The $Q \times I(Q)$ vs. Q axes emphasize the high-angle region. Curves are re-scaled for clarity.

thymine bases do not have favorable stacking interactions [34], no helix-coil transition is expected in this temperature range (in contrast, polypurines adopt a single-stranded helical form at low temperature [35]). The slight increase in R_G for dT_{10} at 4 C may be due to increased density of hydration water around the molecule, which has been shown to have a measurable effect on the scattering from biomolecules [36]. Measurements of the thermodynamics of DNA hairpins with poly(dT) loops of varying length have shown that the configuration entropy of the loop is a major barrier to forming collapsed states [37]. Our temperature-dependent SAXS data suggest that the configuration entropy of poly(dT) loops will also be important at low temperatures.

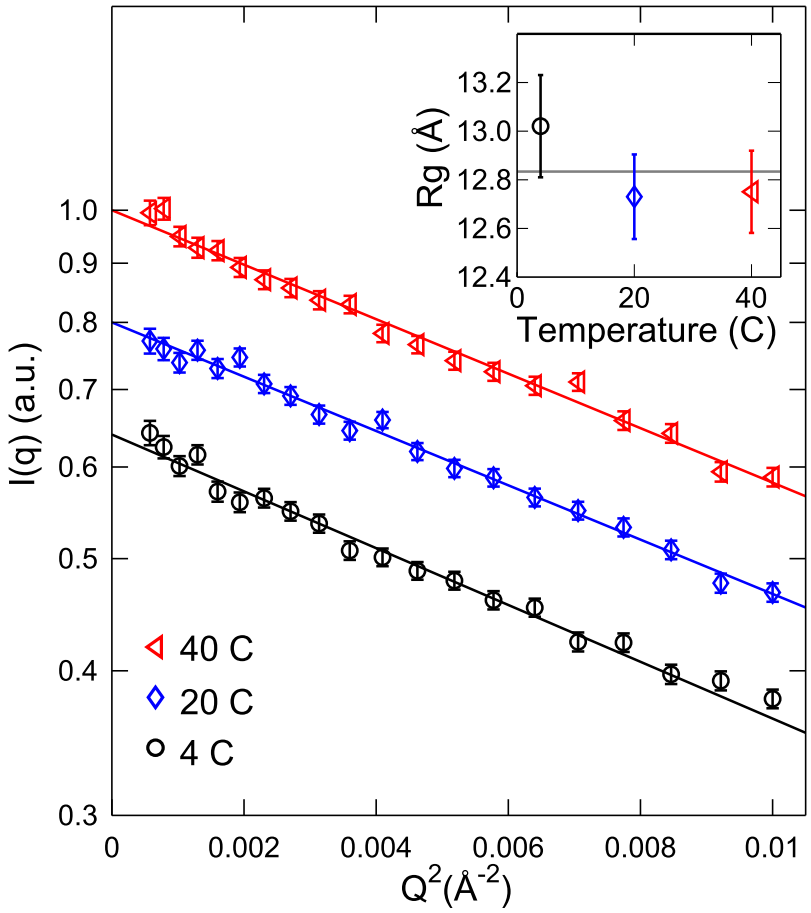

Figure 3.4: SAXS data for dT10 were collected as a function of temperature between 4C and 40C at MacCHESS beamline F2. The data are presented as a Guinier plot (logarithmic intensity vs. q^2), and the curves are offset by a multiplicative factor for clarity. Linear fits are shown as solid lines. The radii of gyration at each temperature were found from the slopes of these lines according to the Guinier law: $I(q) \propto \exp(-q^2 R_G^2/3)$ (inset).

3.2.3 Molecular dynamics simulations of dT_{10} converge to compact states

DT_{10} in 100 mM NaCl at 300K was simulated by three different force fields: (i) OPLSAA with Aqvist ion parameters (described in ref. [20]) (ii) AMBER-99 with updated ion parameters [38] and (iii) AMBER-99 with correction to parameters for the DNA backbone torsion angles α and γ (AMBER-pbsc0) [39]. Each system was simulated for a sufficient length of time to converge both R_G and the ion distributions. The OPLSAA simulation ran for 35ns of Replica

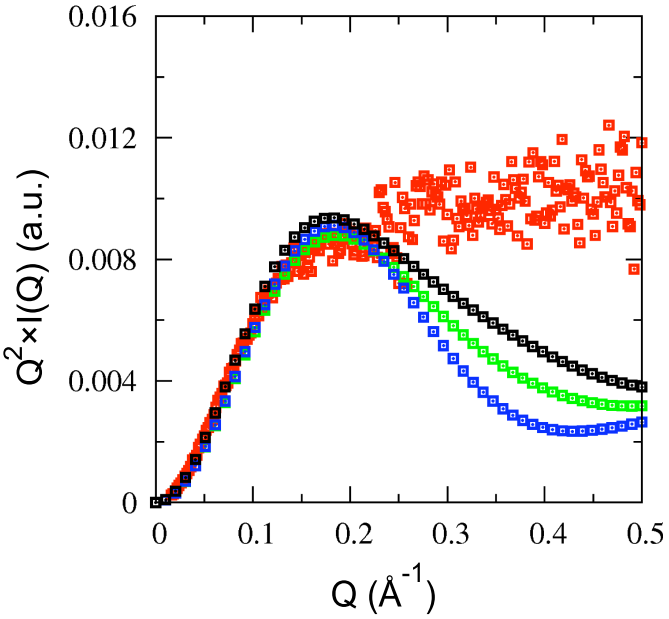

Figure 3.5: MD results are compared with experimental SAXS data (red) in a Kratky plot. The converged results for three different force fields are shown: OPLSAA (blue), AMBER-99 (green) and AMBER-pbsc0 (black).

Exchange, AMBER-99 for 140ns at a single temperature, and AMBER-pbsc0 for 45ns at single temperature. For comparison with the SAXS data, the apparent radii of gyration for each simulation were found from a Guinier fit to the ensemble-averaged scattering curves predicted using CRYSOL [40]. R_G values of 9.4 Å (OPLSAA), 10.0Å (AMBER-99) and 10.5 Å (AMBER-pbsc0) were obtained. These radii of gyration are significantly smaller that the experimental value of 12.8 ± 0.2 Å, regardless of which force field was used. The scattering curves were also compared using a Kratky plot to emphasize the polymer scaling properties in Figure 3.5. The scattering profiles computed from MD ensembles are consistent with compact states relative to the experimental data.

3.3 Conclusions

In MD simulations of ssDNA, attractive interactions overcome steric and electrostatic repulsion, leading to collapsed conformations. In contrast, SAXS data for identical systems show that steric and electrostatic repulsion effects dominate, leading to random coil-like conformations. It is possible that the collapsed conformations, observed here and elsewhere [27], represent a small sub-population that is not observed in ensemble measurements such as SAXS. Previous SAXS experiments carried out at room temperature could not rule out the existence of a small population of collapsed states; however, our data spanning the range of 4-40C makes the existence of a collapsed fraction at room temperature far less likely. It is far more likely that force field parameters are not optimal for predicting single-stranded nucleic acid structure and thermodynamics. We note that previous attempts to benchmark MD force fields using simulations of G-quadruplex loops reached a similar conclusion [12].

Force fields are continually re-tuned when new chemical data or quantum mechanical calculations become available. Currently, the most popular way to validate these force field changes is by de-novo folding, for example of small RNA and protein structures that are already known by NMR or crystallography. It is then assumed that simulations will give accurate dynamical information, i.e. excursions from the known structure and the relative free energies of these states, as well as the rates for transitioning between them. However, we argue that accurate prediction of the conformational ensemble of an intrinsically-disordered molecule would provide an additional, useful validation tool. As a side benefit, accurately simulating the unfolded state should increase success rates for de-novo folding calculations [1].

Poly(dT) is a good candidate for force-field validation because it has a well-studied random coil conformation that persists over a large range of temperatures in physiologically-relevant conditions. Most importantly, all-atom MD simulations using the older force fields fail to re-

produce the most fundamental, average properties of the ensemble as assessed by SAXS, such as the radius of gyration and the polymer scaling properties that manifest at large scattering angles. Since scattering curves can be predicted directly from the MD simulation without any free parameters [41], SAXS is an ideal validation tool. In fact, given the ease of collecting SAXS data at modern synchrotron sources, we would argue that structural ensembles proposed by molecular dynamics should always be presented alongside supporting x-ray scattering data.

In conclusion, it is the force-fields, and not the computational resources, that currently limit the usefulness for MD for the critical challenge of understanding nucleic acid folding and dynamics. Promised applications of MD, such as developing new anti-bacterial agents targeting RNA folding, will require accurate and well-validated force fields. Although Chen & Garcia have corrected aberrant base-stacking energies present in force fields used here [1], these parameters are unlikely to be the only ones needing attention. New simulations currently underway with re-tuned force fields will be bench-marked using the proposed methodology.

Acknowledgments

We thank Arthur Woll at CHESS and Richard Gillilan at MacCHESS for experimental assistance and Alan Chen for insightful discussions. This work was supported by the NIH (R01-GM085062). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0936384, and the MacCHESS resource is supported by the National Institute of General Medical Services (NIGMS) award GM-103485.

References

[1] A. A. Chen and A. E. García, "High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations", Proc. Natl. Acad. Sci. U.S.A. **110**, 16820–16825 (2013).

[2] T. R. Cech, "The ribosome is a ribozyme", Science **289**, 878–879 (2000).

[3] L. Cartegni, S. L. Chew, and A. R. Krainer, "Listening to silence and understanding nonsense: exonic mutations that affect splicing", Nat. Rev. Genet. **3**, 285–298 (2002).

[4] V. N. Kim, "MicroRNA biogenesis: coordinated cropping and dicing", Nat. Rev. Mol. Cell Biol. **6**, 376–385 (2005).

[5] Y. Wan, M. Kertesz, R. C. Spitale, E. Segal, and H. Y. Chang, "Understanding the transcriptome through RNA structure", Nat. Rev. Genet. **12**, 641–655 (2011).

[6] A. Yonath, "Antibiotics targeting ribosomes: resistance, selectivity, synergism, and cellular regulation", Annu. Rev. Biochem. **74**, 649–679 (2005).

[7] A. Serganov and E. Nudler, "A decade of riboswitches", Cell **152**, 17–24 (2013).

[8] A. Schneemann, "The structural and functional role of RNA in icosahedral virus assembly", Annu. Rev. Microbiol. **60**, 51–67 (2006).

[9] J. S. Kieft, "Viral IRES RNA structures and ribosome interactions", Trends Biochem. Sci. **33**, 274–283 (2008).

[10] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, and J. A. Schloss, "The potential and challenges of nanopore sequencing", Nat. Biotechnol. **26**, 1146–1153 (2008).

[11] P. J. Bond, A. T. Guy, A. J. Heron, H. Bayley, and S. Khalid, "Molecular dynamics simulations of dna within a nanopore: arginine–phosphate tethering and a binding/sliding mechanism for translocation", Biochemistry **50**, 3777–3783 (2011).

[12] E. Fadrná, N. Špačková, J. Sarzynska, J. Koca, M. Orozco, T. E. Cheatham III, T. Kulinski, and J. Šponer, "Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields", J. Chem. Theory Comput. **5**, 2514–2530 (2009).

[13] A. Mládek, M. Krepl, D. Svozil, P. Čech, M. Otyepka, P. Banáš, M. Zgarbová, P. Jurečka, and J. Šponer, "Benchmark quantum-chemical calculations on a complete set of rotameric families of the DNA sugar–phosphate backbone and their comparison with modern density functional theory", Phys. Chem. Chem. Phys. **15**, 7295–7310 (2013).

[14] D. E. Draper, "A guide to ions and RNA structure", RNA **10**, 335–343 (2004).

[15] D. E. Draper, "Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water", Biopolymers **99**, 1105–1113 (2013).

[16] M. A. González and J. L. Abascal, "The shear viscosity of rigid water models", J. Chem. Phys. **132**, 096101 (2010).

[17] J. Witz and V. Luzzati, "La structure des acides polyadenylique et polyuridilique en solution: etude par diffusion centrale des rayons X", J. Mol. Biol. **11**, 620–630 (1965).

[18] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA", Proc. Natl. Acad. Sci. U.S.A. **109**, 799–804 (2012).

[19] A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, "Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle X-ray scattering", Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. **86**, 021901+ (2012).

[20] S. P. Meisburger, J. L. Sutton, H. Chen, S. A. Pabit, S. Kirmizialtin, R. Elber, and L. Pollack, "Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: beyond the wormlike chain model", Biopolymers **99**, 1032–1045 (2013).

[21] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle", Recl. Trav. Chim. Pays-Bas **68**, 1106–1122 (1949).

[22] J. Skolnick and M. Fixman, "Electrostatic persistence length of a wormlike polyelectrolyte", Macromolecules **10**, 944–948 (1977).

[23] J. L. Barrat and J. F. Joanny, "Persistence length of polyelectrolyte chains", Europhys. Lett. **24**, 333+ (2007).

[24] T. Odijk, "Polyelectrolytes near the rod limit", J. Polym. Sci. Polym. Phys. Ed. **15**, 477–483 (1977).

[25] B. Y. Ha and D. Thirumalai, "Electrostatic persistence length of a polyelectrolyte chain", Macromolecules **28**, 577–581 (1995).

[26] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck, "Nonlinear low-force elasticity of single-stranded DNA molecules", Phys. Rev. Lett. **102**, 068301 (2009).

[27] J. M. Martínez, S. K. Elmroth, and L. Kloo, "Influence of sodium ions on the dynamics and structure of single-stranded DNA oligomers: a molecular dynamics study", J. Am. Chem. Soc. **123**, 12279–12289 (2001).

[28] K. Chakraborty, S. Mantha, and S. Bandyopadhyay, "Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium", J. Chem. Phys. **139**, 075103 (2013).

[29] S. S. Nielsen, M. Møller, and R. E. Gillilan, "High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell", J. Appl. Crystallogr. **45**, 213–223 (2012).

[30] O. Glatter, "A new method for the evaluation of small-angle scattering data", J. Appl. Crystallogr. **10**, 415–421 (1977).

[31] S. Hansen, "Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering", J. Appl. Crystallogr. **33**, 1415–1421 (2000).

[32] P. J. Flory, *Principles of polymer chemistry* (Cornell University Press, 1953).

[33] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution", Q. Rev. Biophys. **40**, 191–285 (2007).

[34] W. Saenger, *Principles of nucleic acid structure* (Springer-Verlag, 1984).

[35] H. Eisenberg and G. Felsenfeld, "Studies of the temperature-dependent conformation and phase separation of polyriboadenylic acid solutions at neutral ph", J. Mol. Biol. **30**, 17–37 (1967).

[36] D. Svergun, S. Richard, M. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, "Protein hydration in solution: experimental observation by X-ray and neutron scattering", Proc. Natl. Acad. Sci. U.S.A. **95**, 2267–2272 (1998).

[37] S. V. Kuznetsov, C.-C. Ren, S. A. Woodson, and A. Ansari, "Loop dependence of the stability and dynamics of nucleic acid hairpins", Nucleic Acids Res. **36**, 1098–1112 (2008).

[38] I. S. Joung and T. E. Cheatham, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations", J. Phys. Chem. B **112**, 9020–9041 (2008).

[39] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco, "Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers", Biophys. J. **92**, 3817–3829 (2007).

[40] D. Svergun, C. Barberato, and M. H. J. Koch, “CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates”, J. Appl. Crystallogr. **28**, 768–773 (1995).

[41] S. Park, J. P. Bardhan, B. Roux, and L. Makowski, “Simulated X-ray scattering of protein solutions using explicit-solvent models”, J. Chem. Phys. **130**, 134114 (2009).

CHAPTER 4

ORIENTATION CORRELATION FUNCTION OF SINGLE-STRANDED DNA MEASURED USING SMALL-ANGLE X-RAY SCATTERING AND ENSEMBLE OPTIMIZATION

Steve P. Meisburger[1] and Lois Pollack[1]

1. Applied and Engineering Physics, Cornell University, Ithaca, New York

Abstract

The mechanical properties of single-stranded RNA and DNA are fundamental to the biophysics of nucleic acid structural transitions. Understanding the electrostatic contribution to flexibility due to the negatively charged phosphate backbone remains a challenge. One way to parametrize flexibility of a chain is using the orientation correlation function of tangent vectors, a quantity of central importance in polyelectrolyte theory. Previous measurements of single-stranded DNA using techniques sensitive to global structure have been interpreted via the wormlike chain (WLC) model that assumes an exponential decay of the correlation function. However, recent force spectroscopy measurements and theoretical work suggest that the orientation correlation function for single-stranded DNA may be fundamentally different from the WLC prediction.

To investigate the correlation function for single-stranded DNA from a structural perspective, we collect x-ray scattering data for DNA homopolymers under conditions of strong electrostatic repulsion. These data are fit using optimized ensembles of atomic models, and correlation functions are computed directly from the ensembles. We find that the correlation function for polydeoxythymidine (poly(dT)) diverges from exponential behavior for nucleotides separated by more than two neighbors. In contrast, a power-law decay proposed in recent theoretical

work agrees well for nucleotide separations up to ~ 10. To investigate the effect of sequence, we also applied this technique to poly(dA). Ensemble optimization for poly(dA) converges to models with significant helical content and an oscillatory correlation function. Thus, our analysis of the scattering data is consistent with the hypothesized power-law decay of the orientation correlation function for chains similar to poly(dT) that lack significant base stacking. Possible applications of the method for modeling flexible regions of RNA and DNA molecules are also discussed.

4.1 Introduction

The flexible nature of single-stranded DNA and RNA molecules is central to their diverse roles in biology. DNA is a polyelectrolyte with a charge of -1e per nucleotide at neutral pH. This charge adds an important electrostatic component to the flexibility and implies strong interactions with counterions. Understanding the complex relationship between conformation and charge-screening by ions remains a challenge for theory and experiment.

In polyelectrolyte theory, the directional persistence of a chain is described by the orientation (or tangent-vector) correlation function (OCF) [1], illustrated in Figure 4.1. Recent simulations of flexible charged polymers in salt solutions show that the OCF decays as a power law [2], rather than the exponential assumed in earlier theories. Power-law decay of the OCF was also shown to be consistent with anomalous scaling of force extension curves for polynucleotides at high forces [3–6]. However, direct evidence from structural techniques is lacking.

Small angle x-ray scattering (SAXS) is a widely-used technique for studying flexible proteins and nucleic acids [7–11]. In principle, SAXS from single-stranded DNA in solution should be sensitive to the functional form of the OCF. However, previously-used expressions relating the SAXS curve to intrinsic properties of the underlying polymer [12] assume an exponentially-

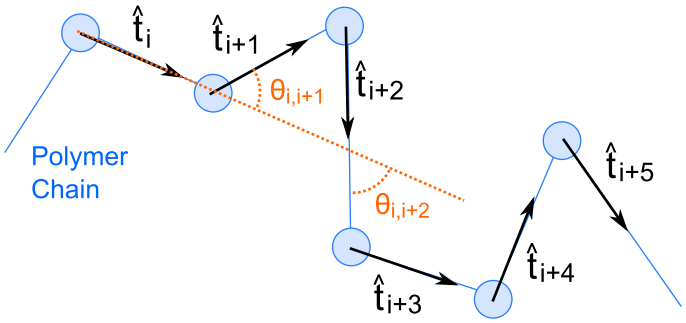

Figure 4.1: The orientation correlation function of a polymer chain is defined for normalized bond vectors $\hat{t}_i$ as $\langle \hat{t}_i \cdot \hat{t}_j \rangle = \langle \cos \theta_{i,j} \rangle$, where the average is computed over chain configurations, with a constant separation of the bond vectors $|i-j|$.

decaying correlation function embodied by the wormlike chain model [13]. Here, we describe a strategy that is inspired by similar studies of disordered proteins [14] for fitting x-ray data with optimized ensembles of nucleic acid models. Using this ensemble optimization method (EOM) to fit the data using realistic structural pools, we compute correlation functions directly from the atomic coordinates of the selected models.

While SAXS is low-resolution and insensitive to details on the scale of atomic bonds, in single-stranded nucleic acids the backbone conformation (e.g. base-stacking [15, 16]) may affect global structure on the length-scales observable by SAXS (~ 10Å). Previous approaches for generating candidate nucleic acid models for selection by SAXS data include coarse-grained simulation [17] and assembly of fragments from crystal structures [18]. Pools generated in this way have been used for identifying a single most probable structure [19–22] and for ensemble optimization methods (EOM) [23]. However, these modeling strategies were intended for RNA molecules that form well-defined secondary and tertiary structures, and are unlikely to be appropriate for modeling unstructured nucleic acids. Although molecular dynamics offers a possible alternative, large simulation boxes and long equilibration times required for single-stranded polynucleotides make MD computationally expensive. Furthermore, the torsion-angle potentials used in modern force fields are not necessarily reliable: they have been unable to reproduce single-stranded DNA motifs within larger structures (e.g. [24]).

Therefore, we adopt a modeling approach for unstructured DNA based on the long-standing observation that nucleic acid backbone torsion angles tend to be correlated with one another between adjacent nucleotides [25–31]: in RNA, 46 distinct torsion angle sets account for the majority of conformations in high-resolution structures [32]. Tools for model-building based on these discrete structure libraries have been developed [33, 34]. While DNA is more conformationally plastic than RNA (the 2'-OH present in RNA, but not in DNA, is thought to stabilize particular conformations), the DNA backbone conformers nonetheless cluster into distinct categories [35, 36].

To generate pools of DNA structures, we build the chain from a series of discrete models, drawn from a small library of representative examples. Each structure in the library is defined from sugar-to-sugar in a unit known as a suite[27, 32]. Each suite definition includes seven backbone torsion angles $\delta(i-1)\varepsilon\xi\alpha\beta\gamma\delta$ and two glycosidic torsions $\chi(i-1)$ and χ, illustrated in Figure 4.2. When building chains, the suites overlap at their 3' and 5' ends. Therefore, we use two values for δ and χ angles shared among suites: δ is assigned canonical values for either $C2' - endo$ or $C3' - endo$ sugar pucker, while χ may be either *anti* or *syn* [37]. Chains were additionally free of steric clashes, defined as the significant overlap of Van der Waals radii of non-bonded atoms [33, 38].

Each suite is assigned a statistical weight, equal to the probability for it to occur as an isolated dinucleotide. This probability is in general unknown, but may be constrained by available data. As in previous work, the statistical weights are collectively re-scaled in order to obtain the correct sequence-dependent equilibrium constant for sugar pucker measured by NMR [34]. Furthermore, weights for different χ conformations are assigned according to the known sequence preferences: for purines, *syn* and *anti* are given equal prior weight, while pyrimidines are always *anti*. Note that the *syn* conformation is thought to contribute entropically to the cooperativity of base-stacking [39], an effect we seek to model correctly. Finally, each set of

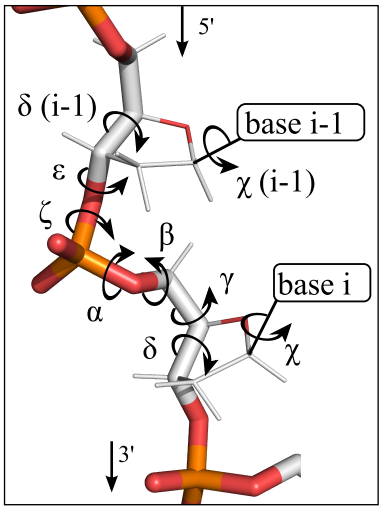

Figure 4.2: Definition of backbone torsion angles for a dinucleotide suite, following Refs. [27, 32].

five backbone torsion angles $\varepsilon\xi\alpha\beta\gamma$ is initially given a uniform weight that is later refined according to the apparent selection pressure for particular conformers over multiple rounds of ensemble optimization. In this way, significant torsion angle bias is avoided in the structure pools, beyond the limitations imposed by the discrete library. In order for this strategy to converge, the number of parameters in the optimization is kept reasonably small (~ 10) by limiting the number of discrete torsion angle sets to a similar number.

4.2 Materials and methods

4.2.1 X-ray scattering from DNA homopolymers in solution

Synthetic HPLC-purified DNA oligomers dT_{30} and dA_{30} were purchased from Integrated DNA Technologies (Coralville IA). Lyophilized powders were re-suspended and exchanged 4 times with buffer containing 20 mM NaCl, 1 mM Na+MOPS pH 7.0 using Amicon Ultra-0.5, 3 kDa cutoff centrifugal concentrators (EMD Millipore, Billerica, MA). SAXS profiles were measured

for DNA concentrations of 200, 100, and 50 μM at CHESS beamline G1 as described previously [40]. Buffer-subtracted curves were extrapolated linearly to the zero concentration to correct for interparticle interference effect seen at $q < 0.05\text{Å}^{-1}$ [40, 41]. The extrapolated curves were scaled to match the high concentration curves in the region $0.05 < q < 0.1\,\text{Å}^{-1}$, and the dilute and high concentration curves were stitched together at $q = 0.05\text{Å}^{-1}$. The SAXS profiles and their experimental uncertainties were normalized by the forward scattering $I(q \to 0)$ determined from the indirect Fourier transform [42]. All data analysis was performed using Matlab (MathWorks, Natick, MA) code written in-house.

4.2.2 Geometric parameters of dinucleotides

A set of twelve different $\varepsilon\xi\alpha\beta\gamma$ torsion angle combinations was chosen, with the goal of representing important conformers observed in DNA crystal structures (such as BI and BII stacks) while also sampling extended conformers that are rare in crystal structures, but may be populated in the unfolded state. Because non-helical conformations of DNA are not common in crystals, conformation classes for non-stacked DNA were derived from RNA rotamers. Using these RNA rotamers to model DNA is a reasonable approximation because DNA has fewer steric restrictions than RNA and presumably RNA conformations represent a subset of the possible backbone conformations available to DNA: for example, RNA cannot adopt B-form helical stacks because of steric clashes, while DNA has been observed in both A and B-form helices.

Each torsion angle set was assigned a 2-3 character mnemonic in Table 4.1. The stacked conformations are named according to their class, i.e. "b1" is the BI class [35]. For unstacked conformations, the first two characters of the mnemonic refer to different ranges of the ξ/α pair using the nomenclature in Ref. [25], and the third character specifies the range for γ (p = gauche+, m = gauche-, and t = trans). In RNA, the precise values of $\varepsilon\xi\alpha\beta\gamma$ within a

single conformation range are correlated with $\delta(i-1)$ and δ. However, for simplicity we use a single torsion angle set for all allowed values of δ. The values for $\varepsilon\xi\alpha\beta\gamma$ were taken as the average for RNA suites related by those torsion angles, but having different $\delta(i-1)$ and δ. This is not a severe approximation, as the dinucleotides generated using these average torsion angles are identified as the correct RNA suites by the program suitename [32]. For the unstacked conformations, all values of $\delta(i-1)$, δ, $\chi(i-1)$, and χ are modeled, subject to steric contraints and sequence preferences. For the stacked conformations, a single set of $\delta(i-1)$ and δ is allowed, and χ is always *anti*.

For determining steric clashes, atoms were assigned Van der Waals radii of 1.52 (O), 1.7 (C), 1.8 (P), and 1.55 Å (N) as in previous discrete models for nucleic acids [33]. A steric clash was defined as a Van der Waals overlap greater than 0.42 for non-bonded atoms (this value was relaxed slightly from the overlap cutoff of 0.4 defined in the nucleic acid model validation tool MolProbity [38] to accommodate imperfect stacking geometry resulting from common values for δ and χ for all suites). Bond angles and distances were taken from the XPLOR high-resolution parameter set [43]. When building chains using a sequence of suites, the γ angle of the 5'-terminal sugar is undefined, and was therefore assigned the canonical value of 52.5°. Base torsion angles χ for the *anti* conformation of purine and pyrimidine bases were adjusted slightly within their allowed ranges to prevent steric clashes in stacked the dTpdT and dApdA suites, and are reported in Table 4.2.

The model-building technique we describe is very general, and may be applied to RNA and DNA molecules of arbitrary sequence. However, for the purposes of this paper, we implemented only DNA homopolymer models with A and T bases. The sugar-pucker equilibrium constants for dTpdT and dApdA dinucleotides are 1.9 and 24, respectively [34, 44, 45]. For poly(dA), such a large equilibrium constant implies that C3'-endo sugar pucker will be exceedingly rare (less than 2%). Therefore, for simplicity suites with C3'-endo sugars were eliminated from the

Mnemonic	Derived from	ε	ξ	α	β	γ	Restrictions on $\{\delta(i-1), \delta\}$
a1p	5d,6d,6p	234	81	65	159	53	none
a1t	5j,5q,6j	224	72	67	115	178	none
a2p	4p,4d,7p,3d	248	206	72	197	57	none
p1p	9a,0b,0a	227	124	283	160	51	none
p3p	2a,2I,1I,1m	239	290	292	212	55	none
p3t	2h	261	290	296	177	176	none
s3p	1g,1z	213	285	182	161	51	none
s3t	1t	199	289	180	195	178	none
b1	BI	184	262	302	179	45	{2,2}
b2	BII	245	172	297	142	46	{2,2}
a1	AI	205	285	294	172	55	{3,3}
b2a	BII-AI	257	186	60	224	196	{3,2}

Table 4.1: Twelve backbone torsion angle sets that span adjacent sugar rings, {ε, ξ, α, β, γ} were derived from RNA backbone rotamers (lower-case rotamer names from Ref. [32]) or DNA classes (upper-case, from Ref. [35]). Angles are given in degrees, using the same sign convention as in Ref. [35]. The torsion angle δ was assigned canonical values according to the sugar pucker of either $C2' - endo$ or $C3' - endo$, denoted by numerals 2 and 3. Numerical values for δ and χ are given in Table 4.2.

Sugar-pucker	δ	χ(anti, pyrimidines)	χ (anti, purines)	χ(syn, purines)
C2'-endo	145.2	259	250.3	70.0
C3'-endo	81.1	204	200.5	70.0

Table 4.2: Sugar (δ) and base (χ) torsion angles in degrees, defined as in Ref. [35], for the two sugar-pucker types modeled here.

poly(dA) pool. Thus, for dTpdT dinucleotides, there were 40 allowed torsion angle combinations, of which 35 were clash-free. For dApdA, 32 conformations out of 38 were clash-free. Dinucleotide suites with steric clashes were eliminated from the libraries.

4.2.3 Chain-generating algorithm with statistically weighted conformers

In our chain generator, a set of M suites specifies all possible dinucleotide conformations. Each suite is assigned a statistical weight, and the set of weights is stored in a vector w of length M. These weights are defined as the probability for each suite to occur as an isolated dinucleotide.

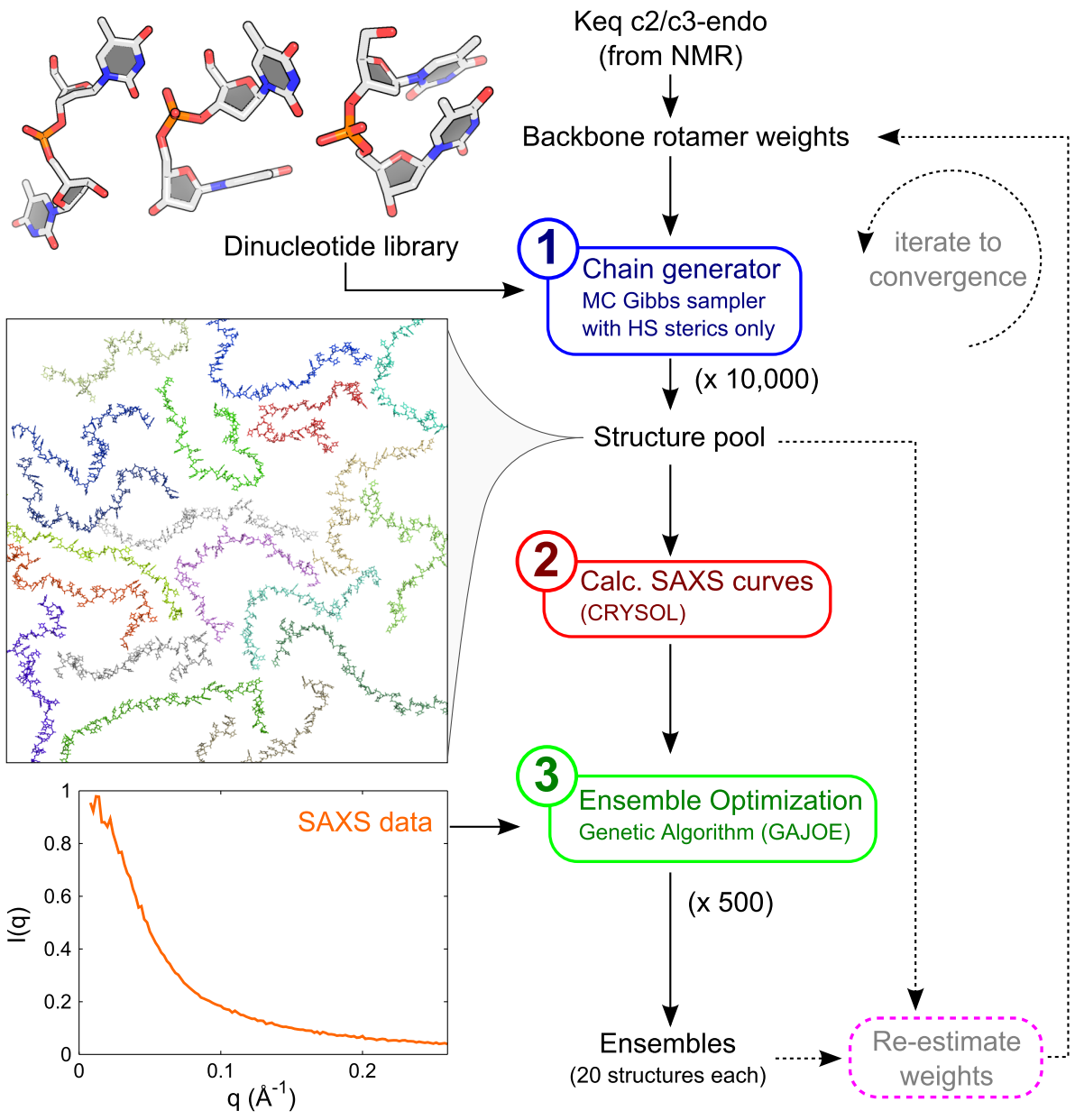

Figure 4.3: Overview of the nucleic acid ensemble optimization procedure.

A conformation of a nucleic acid chain of length N (i.e. N bases and $N-1$ phosphates) is specified by a list c of $N-1$ integers between 1 and M. Because the suite definition includes both the 5' and 3' sugars, adjacent suites overlap. Therefore, the integers obey adjacency rules (a 5'c3 suite may not follow a 3'c2 suite, and a suite with 5'-syn base may not follow a 3'-anti, etc.). Furthermore, steric clashes severely limit the space of possible conformations. To accelerate the computation of steric clashes, all possible chains of length $N = 5$ are pre-computed. For each of these chains, adjacency rule violations and steric clashes are identified. These outcomes are stored in a 4-dimensional ($M \times M \times M \times M$) logical matrix L_4 for fast lookup.

Next, a Monte-Carlo sampling technique generates random chains whose suites are drawn probabilistically according to their weights, while avoiding steric clashes and adjacency rule violations. This method consists of 4 steps:

1. Initialize the conformation list c_0 to a random state that satisfies L_4.
2. Starting from c_{i-1}, use L_4 and w to propose a new state c_i.
3. Compute the 3D model for c_i and search for steric clashes. If clash-free, move to step 4. Otherwise, return to step 2.
4. Save the 3D model for c_i. If the desired number models have been saved, end. Otherwise, set $i = i+1$ and return to step 2.

The implementation of each of these steps is briefly described. In step 1, an initial conformation vector c_0 is set at random. To make c_0 consistent with L_4, changes to single positions in c_0 are proposed randomly, and accepted if the total number of violations, V_{tot} decreases or stays the same. V_{tot} is computed by scanning L_4 across c:

$$V_{tot} = \sum_{j=1}^{N-4} L_4\left(c_0(j), c_0(j+1), c_0(j+2), c_0(j+3)\right)$$

This procedure is repeated until the number of violations is zero (typically a few hundred iterations for $N = 30$).

Step 2 consists of multiple iterations of a Gibbs sampler. In each iteration, an integer n between 1 and $N-2$ is chosen at random. The sampler generates new values for positions n and $n+1$ in c based on the adjacent values and the statistical weights. First, all pairs allowed by L_4 are enumerated. The statistical weight for each pair is computed from the product of the weights for the two suites. The new pair is chosen at random with a probability equal to its statistical weight divided by the sum of the weights for all possible pairs (this is the Gibbs sampling step [46]). The sampler is iterated to generate a new trial state that is sufficiently different from c (we use 50 iterations for $N = 30$).

In step 3, the 3D model is computed from c_i using the geometric rules for nucleic acid bond angles and distances. Because c_i is guaranteed to have no steric clashes between nucleotides separated by less than 5 bases, and because nucleotides have a known maximum size, it is not necessary to compute the pair-wise distance between all atoms. First, pair-wise distance between all C1' atoms are computed. Only pairs of nucleotides separated by more than 5 bases along the chain, and whose C1' atoms are within a cutoff distance of 20 Å [33], are checked for clashes.

In step 4, after a burn-in period of 10 iterations, the 3D models are saved in PDB format. The algorithm stops when the specified number of models has been generated. Steps 1-4 were implemented in MATLAB and optimized for efficient execution on a desktop computer.

4.2.4 Ensemble optimization

SAXS curves are predicted for each model in the pool using CRYSOL [47] with maximum harmonic order of 15, Fibonacci grid of order 18, and default hydration parameters. In order to fit the experimental data, an optimized ensemble of 20 structures is chosen by by running a genetic algorithm for 50 generations using the program GAJOE13 [9]. The selection is repeated

independently in order to accumulate statistics: 50 times for each cycle of EOM used in torsion weight refinement, and 500 times in the final run.

4.2.5 Iterative refinement of the conformer weights

In the calculation of suite weights, each is first assigned a weight $t^{(i)}$ according to its torsion angle set, and then all suite weights are re-scaled to give the correct equilibrium constants for sugar-pucker measured by NMR, as described in Ref. [34]. This process is expressed as $\mathbf{w} = \mathbf{w}(\mathbf{t}, K_{eq}^{C2-endo})$ where $\mathbf{w}(\mathbf{t}, K_{eq})$ is the vector-valued function that assigns weights based on $\mathbf{t}$, and the sugar-pucker equilibrium constant $K_{eq}^{C2-endo}$. During the refinement step, the frequency of each suite in the pool and in the ensembles is calculated (h_{pool} and h_{ens} respectively). A new estimate $\mathbf{t}_{new}$ is found that minimizes the discrepancy between the observed h_{ens} and an expected value for h_{ens} assuming that the frequencies are proportional to the underlying weights:

$$\chi_h^2 = \sum_i \left(h_{ens}^{(i)} - \frac{h_{pool}^{(i)} w^{(i)} \left(\mathbf{t}_{new}, K_{eq}^{C2-endo}\right) / w_{old}^{(i)}}{\sum_j h_{pool}^{(j)} w^{(j)} \left(\mathbf{t}_{new}, K_{eq}^{C2-endo}\right) / w_{old}^{(j)}} \right)^2$$

where w_{old} is the vector of suite weights used by the chain generator to create the pool. A vector $\mathbf{t}_{new}$ that minimizes χ_h^2 is found using the lsqnonlin function in Matlab.

4.2.6 Metrics for ensemble quality and convergence

The selection results are analyzed in terms of the similarity between the starting pool and the ensembles. A convenient way to quantify selection is using the concept of entropy. The Shannon entropy is defined for a probability distribution p of a finite set of states:

$$H(\mathbf{p}) = -\sum_i p_i \ln(p_i) \tag{4.1}$$

If the number of times a model j appears in the ensemble is m_j, and $M_{ens} = \sum_j m_j$ is the total number of models in the ensemble, then the entropy of this distribution may be defined as

$$S_{ens} = H(\mathbf{m}/M_{ens})$$

S_{ens} is zero if only one model appears, and is maximized when all models are equally-represented. However, because the selection algorithm is randomized, the expected value of S_{ens} will deviate from the ideal value due to random chance. Thus, we compare S_{ens} to the expected value for equally probable models, which we estimate by assuming m_j to be Poisson-distributed with a mean of $\lambda = N_{ens}/N_{pool}$,

$$S_0 = -\sum_{k=1}^{\infty} \left(\frac{\lambda^k e^{-\lambda}}{k!} \right) k \ln \left(\frac{k}{N_{ens}} \right)$$

To evaluate the dissimilarity between the pool and ensembles in terms of a selection pressure for specific suites, the frequencies of each suite in the pool and the ensemble are compared. A convenient information-theoretic measure of the dissimilarity of two probability distributions is the Jensen Shannon Divergence (JSD) [48] :

$$JSD(p_1, p_2) = H\left(\frac{1}{2}p_1 + \frac{1}{2}p_2\right) - \frac{1}{2}H(p_1) - \frac{1}{2}H(p_2) \tag{4.2}$$

where $H(p)$ is the Shannon entropy (Equation 4.1).

Finally, the goodness of fit is assessed by the χ^2 value comparing the calculated SAXS curve $I_{calc}(q)$ to the experimental one $I_{exp}(q)$ [9]:

$$\chi^2 = \frac{1}{K-1} \sum_{i=1}^{K} \left(\frac{I_{exp}(q_i) - c\, I_{calc}(q_i)}{\sigma_{exp}(q_i)} \right)^2$$

where c is a scale factor for the calculated data and σ_{exp} is the experimental error estimate. The average χ^2 is calculated from multiple runs of the genetic algorithm

$$\langle \chi^2_{fit} \rangle = N^{-1} \sum_{j=1}^{N} \chi^2_j$$

4.3 Results and Discussion

Small angle x-ray scattering profiles were measured for synthetic 30-mers of deoxythymidine (dT_{30}) and deoxyadenosine (dA_{30}) in 20 mM NaCl, buffered to pH 7.0. These low-salt conditions were chosen to enhance the electrostatic repulsion between phosphates in order to observe the conformation in the swollen state [49, 50]. Because of the low-salt buffer, interparticle interference effects are significant [40, 41]. Therefore, a series of DNA concentrations was measured, and the scattering curves were extrapolated to the dilute limit as described previously for dT_{30} [40]. The extrapolated data are shown in Figure 4.6a, and further details may be found in Methods.

Next, ensemble optimization was performed separately for each homopolymer. The details are described in Methods. Briefly, an initial pool of 1,000 structures was generated using uniformly-weighted torsion angle sets. Ensembles of 20 structures that together fit the SAXS data were identified from the pool using a genetic algorithm. This procedure was repeated 50 times for a total ensemble size of 1,000. Certain torsion angle sets were enriched in the ensembles relative to the starting pool. Therefore, a new pool was generated with re-weighted torsion angles, and the process was repeated 18 times. In the final iteration, a much larger pool of 10,000 structures was fit using 500 runs of the genetic algorithm for calculating the correlation function.

To verify that the iterative re-weighting has the desired effect of increasing the quality of the pool, we monitored several indicators during refinement. The refinement procedure is designed to minimize the difference between suite frequencies in the pool (h_{pool}) and ensembles (h_{ens}). This difference was quantified using the JSD between h_{pool} and h_{ens} (Equation 4.2) calculated at each iteration. As expected, a precipitous decrease in the JSD is observed over the first several iterations, followed by a gradual decline and eventual plateau after about 10 iterations (Figure

4.4a). If the pools improve with each iteration, the apparent selectivity of EOM should decrease, and the ensembles should appear more random (i.e., show less preference for particular models). The selection entropy ($S_{ens} - S_0$) reports on the randomness of the ensembles. As the JSD decreases, ($S_{ens} - S_0$) increases, showing that the refinement procedure indeed generates pools that have greater redundancy (Figure 4.4b). Finally, the ability of EOM to fit the data is measured using the average χ^2 obtained by independent ensemble selection runs. The fits are poor in the first iteration, but they improve rapidly as the torsion angle weights are refined (Figure 4.4b). Finally, convergence was verified by comparing the torsion angle frequencies in the final ensemble and pool of 10,000 structures (Figure 4.5).

From the torsion angles represented in the final pool, some conclusions about the overall conformation of each polynucleotide may be drawn. These torsion angle frequencies are shown in Figure 4.5a. First, poly(dA) shows a strong preference for stacked dinucleotides, which make up ~70% of the total. In contrast, poly(dT) homopolymers do not show a strong preference for stacked relative to unstacked conformations. Although each torsion angle was weighted equally in the initial pool, by the final pool several torsion angles are represented at a level below one per chain, and are therefore unlikely to be essential for fitting the SAXS data. For example, the unstacked conformations with mnemonics a1p, a2p, and s3p defined in Table 4.1 are rarely found in poly(dT) and poly(dA) models, and the A-form stack (labeled a1) is not frequently observed in poly(dT) models.

The torsion angle ranges, weighted by the frequencies with which they are observed in the final pools, are shown for in Figure 4.5b and 4.5c for dT_{30} and dA_{30}, respectively. Remarkably, these torsion angle frequencies reflect the known preferences of the polynucleotide backbone for ξ gauche-minus / trans, α gauche-minus, and γ gauche-plus that characterize the canonical helices. In poly(dT) models, the most common unstacked conformations were p3p, p3t, and s3t, which share the common feature of having the ξ torsion angle of gauche-minus, but with

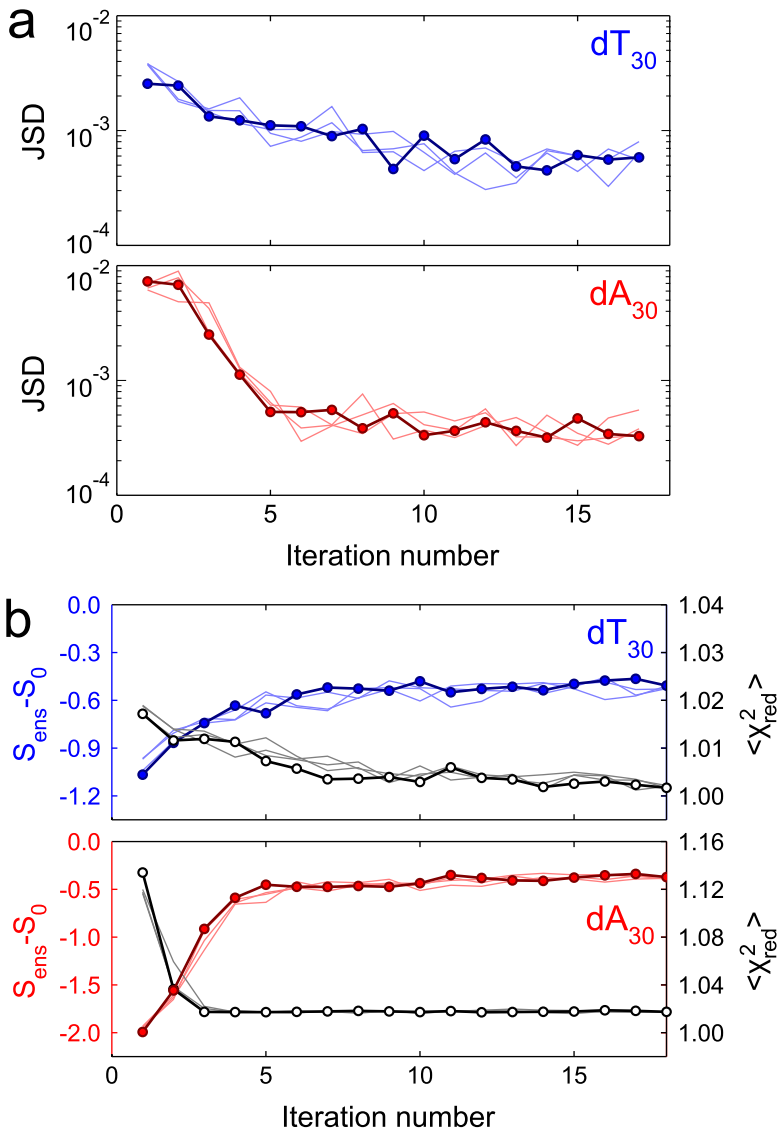

Figure 4.4: Convergence of EOM quality indicators after iterative refinement of torsion angle weights. Thick lines with symbols correspond to the first run of iterative refinement, and thin lines show three independent runs (results in Supplementary Figure A.1). (a) The Jensen Shannon Divergence (JSD) between suite frequency distributions in the pool and ensembles is computed after each iteration of torsion angle weight refinement. The JSD is plotted on a logarithmic scale vs. iteration number for refinements performed separately using the dT_{30} and dA_{30} data and dinucleotide libraries. (b) EOM quality metrics $S_{ens} - S_0$ (selection entropy) and $\left\langle \chi^2_{fit} \right\rangle$ (average reduced χ^2 for the fit of ensembles to data), are shown for dT_{30} (top) and dA_{30} (bottom) as a function of iteration number.

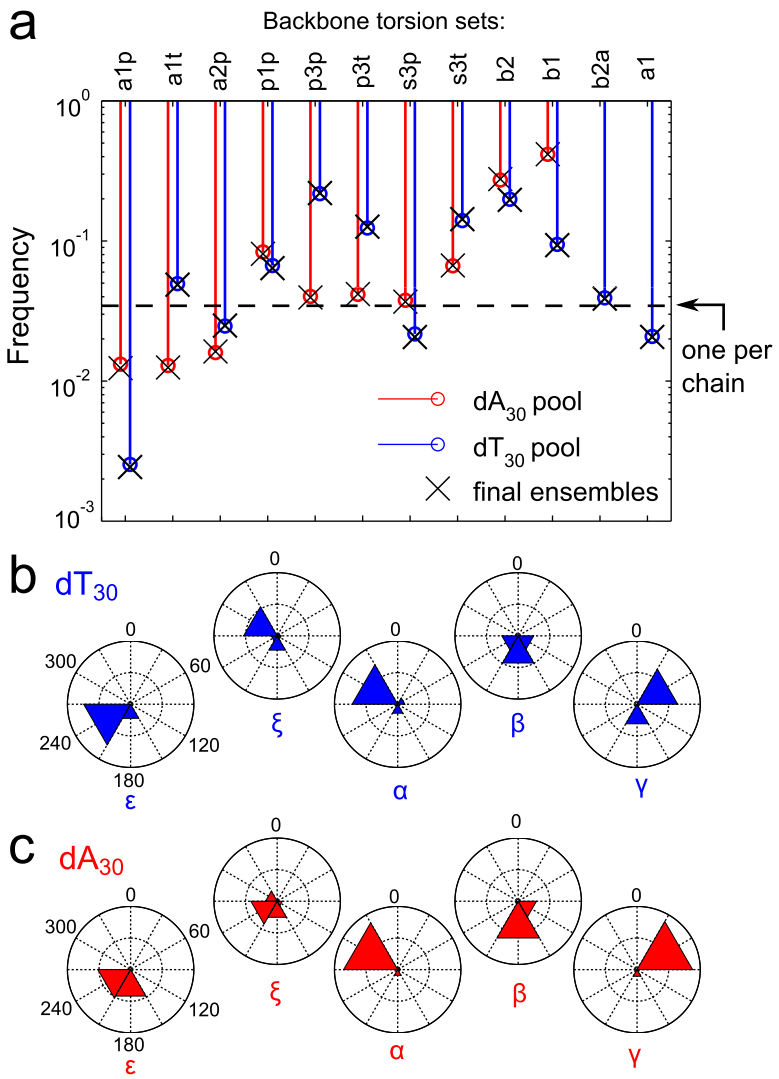

Figure 4.5: Backbone torsion angle frequencies in the optimized ensembles. (a) Frequency of torsion angle sets in the final pool (circles) and ensembles (crosses) are compared for dT_{30} and dA_{30} after 18 rounds of refinement. Labels for torsion angle sets are defined in Table 4.1. Similar frequencies in the pool and ensembles indicate convergence with respect to selective pressure for particular torsion angles. Poly(dA) shows a strong preference for the stacked BI and BII conformations, while poly(dT) contains a mixture of stacked bases and the unstacked p3p, p3t, and s3t conformations. Several torsion angle sets appear only rarely, at a frequency of less than one per chain (dashed line), including a1p, a2p and a1. Four independent runs of torsion angle refinement starting from uniform torsion angle weights demonstrate the uniqueness of the solution: see Supplementary Figure A.1. (b,c) Conformation ranges of the backbone torsion angles in the final ensemble are shown as a rose plot, where the radial extent of the polygon shows the fraction of torsion angles that lie within each 60-degree angular range (outer-ring = 1.0, inner ring = 0.5). The clockwise angle convention is shown explicitly for ε, and is same for all plots.

γ either gauche-plus or trans, and with BI-like or BII-like values of ε and β. Thus, from the perspective of the individual torsion angle ranges, the conformations of poly(dA) and poly(dT) do not appear radically different from "normal" DNA.

Although the final round of selection does not show any strong preference torsion angle frequency (the JSD is ~zero), the selection of particular models is still far from random ($S_{ens} < S_0$). In generating pools, only short-ranged steric clashes are considered; however, electrostatic interactions in the low-salt solution are long-ranged. Thus, although torsion angle weights set the local backbone stiffness (i.e. persistence length) due to electrostatic repulsion and other effects, the remaining long-ranged correlations must still be selected by EOM. One way this effect can be seen is by comparing the average radii of gyration for the pools vs. the ensembles. The poly(dA) models have an RMS average radius of gyration (Rg) of 28.8 Å for the pool, which is less than the ensemble average value of 30.6Å. Similarly, ensemble selection for for poly(dT) increases the R_G from 31.5 Å to 33.5 Å. See also Supplemental Figure A.2.

Finally, we compute the orientation correlation function from the ensembles. The backbone tangent is defined as a normalized vector $\hat{t}$ along the fictitious bond spanning adjacent phosphorus atoms. The dot product of tangent vectors for all i, j pairs of nucleotides are computed, excluding 5 nucleotides at the 3' end and 5 nucleotides at the 5' end in order to suppress end effects. The average of these dot products, $\langle \hat{t}_i \cdot \hat{t}_j \rangle = \langle \cos\theta_{i,j} \rangle$ is computed for equal values of the separation, $|i-j|$ (for an illustration, see Figure 4.1). The orientation correlation functions (OCF) for poly(dA) and poly(dT) are shown in Figure 4.6b.

For poly(dT), which has been described as a model polyelectrolyte (e.g. [3, 51]), we fit two different polyelectrolyte theories directly to the OCF. The wormlike chain model of the form

$$\langle \cos\theta_{i,j} \rangle = e^{-|i-j|/n_p} \tag{4.3}$$

agrees with the OCF for values of the separation up to $|i-j| = 2$ with a persistence number $n_p = 2.5$. When multiplied by the average inter-phosphate distance in the ensembles, n_p im-

plies a persistence length of $\sim 17\,\text{Å}$. For backbone tangent vectors separated by more than 2 nucleotides, the worm-like chain model underestimates the orientational correlations. Similarly, polyelectrolyte simulations show a crossover in the orientational correlation function from semiflexible chain-like behavior to a long-ranged excluded volume-type interaction, where the decay is more gradual (e.g. [1, 2]). Therefore, we also fit the power-law *ansatz* from Ref. [2],

$$\left\langle \cos\theta_{i,j} \right\rangle = C\,|i-j|^{-\gamma} \tag{4.4}$$

The best-fit values of the parameters are $C = 0.322$ and $\gamma = 0.58$, and the power-law behavior is a good fit for separations up to $|i-j| \sim 10$. Polyelectrolyte simulations of a comparatively long, ssDNA-like polymer with in 20 mM ionic strength solution found $C \approx 0.55$ and $\gamma \approx 0.4$ [2]. Agreement with theory is not perfect. However, considering the approximations used in the polyelectrolyte simulations and the difference in chain lengths, we conclude that the general description of poly(dT) as a polyelectrolyte with a power-law OCF is correct.

The OCF was also computed for poly(dA) ensembles. Compared with poly(dT) in Figure 4.6b, the OCF for poly(dA) strongly deviates from either wormlike chain or power-law behavior. The reason for the oscillatory behavior of the OCF is the preference of base-stacked conformations for right-handed helical curvature. Several example models from the final pools of poly(dA) and poly(dT) are shown in Figure 4.6a for visual comparison. In these models, regular undulation of the poly(dA) backbone is apparent, while poly(dT) appears more or less isotropically flexible. For poly(dA), the OCF initially decays quickly due to the intrinsic backbone curvature of base stacks, and returns eventually to a local maximum around one helical period. The value of $\left\langle \cos\theta_{i,j} \right\rangle$ at this local maximum reflects the overall persistence of these helices due to occasional breaks in the base-stacking pattern. Remarkably, the OCF for poly(dA) at $|i-j| \sim 8$ is very similar to the OCF for poly(dT). Apparently, in 20 mM NaCl, the chain stiffness provided by base-stacking in poly(dA) is only marginally greater than the electrostatic stiffness due to screened repulsion between chain segments.

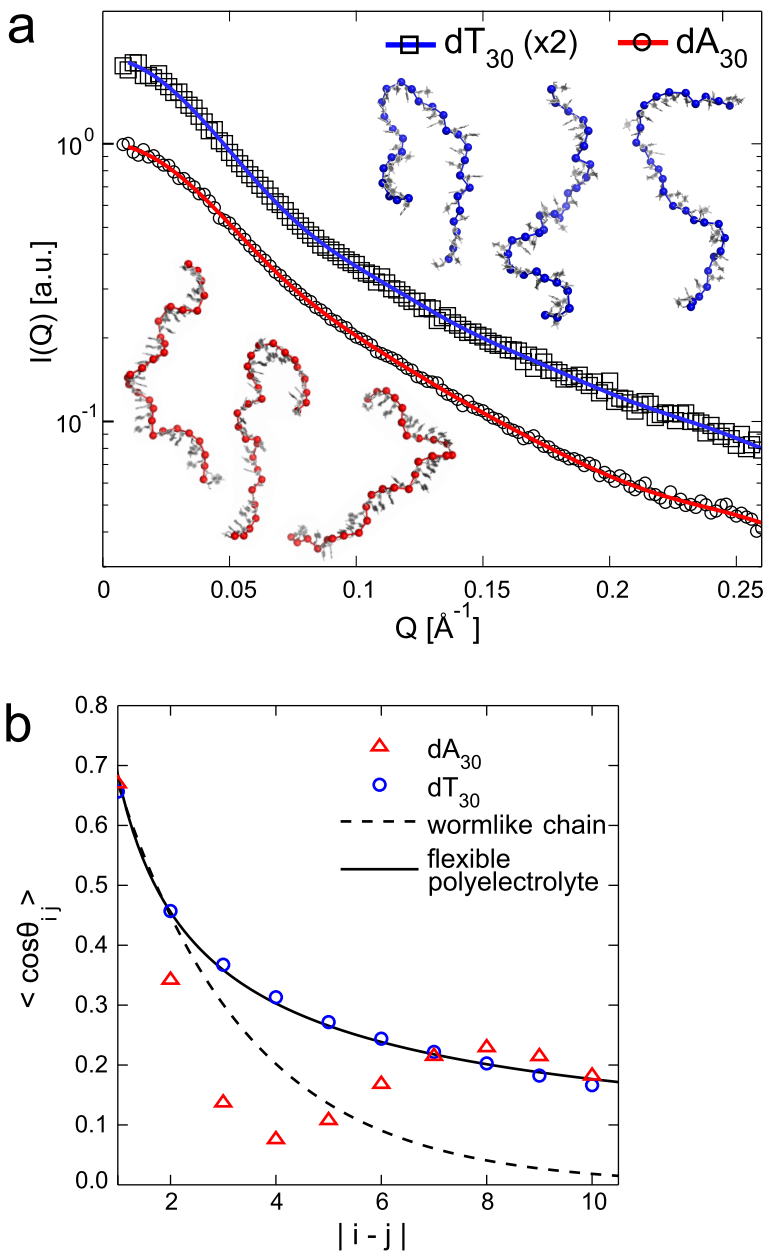

Figure 4.6: (a) SAXS profiles $I(q)$ vs. q ($q = 4\pi\sin(2\theta/2)/\lambda$ is the momentum transfer) for DNA homopolymers dT_{30} and dA_{30} in 20 mM NaCl and neutral pH. Profiles were measured at three DNA concentrations and extrapolated to the dilute limit. Profiles are normalized by $I(q \to 0)$, and the dT_{30} profiles are multiplied by 2 for clarity. The data are shown as symbols, and the fit from EOM as a solid line. Example chains from the final ensembles are shown for dT_{30} (blue) and dA_{30} (red). (b) The orientation correlation function of pseudo-bonds spanning the phosphorus atoms of adjacent nucleotides was computed from the EOM ensembles (for repeatability, see Supplemental Figure A.2). The dT_{30} correlation function is shown with two different polymer models; the wormlike chain (Equation 4.3) with persistence number $n_p = 2.5$; and the flexible polyelectrolyte model of Ref. [2] (Equation 4.4) with $c = 0.322$ and $\gamma = 0.58$. Oscillation in the correlation function for dA_{30} is caused by persistent local curvature (i.e. single-stranded helices).

4.4 Conclusions

The theoretically proposed correspondence between high force scaling and the orientation correlation function in zero force [2] is confirmed experimentally for ssDNA by data presented here in combination with force spectroscopy measurements of similar systems [5]. The use of explicit, atomic models of the DNA avoids the potential ambiguity in earlier interpretations of SAXS data where the chain segments were modeled as uniform cylinders [12, 40, 49, 50]. Whereas earlier SAXS studies modeled ssDNA as an isotropically flexible polymer [40, 49, 50], the generality SAXS-EOM technique allows us to investigate the effects of base-stacking interactions that prefer right-handed helical curvature. We find that the solution structure of dA_{30} is characterized by short sections of rigidly base-stacked nucleotides with an overall conformation ensemble that appears highly flexible. This model for poly(dA) is consistent with its mechanical properties observed recently by force spectroscopy [52].

Since SAXS is a low resolution technique, many of the conclusions about specific populations of the backbone torsions await confirmation by high resolution methods. We successfully integrated a limited amount of NMR data constraining the δ torsion angle according to known sugar pucker equilibrium constants. In the future, combining the SAXS-EOM techniques with NMR measurements in a more comprehensive way should lead to more realistic backbone models. Furthermore, all-atom molecular dynamics simulations of short segments of DNA could be used to build realistic conformations of longer chains for EOM.

Future studies will extend this work to model RNA homopolymers poly(rU) and poly(rA), which share many of the same properties as their DNA analogs, but have shown distinct differences in X-ray scattering, single-molecule FRET [49], and force spectroscopy measurements [6] that are not fully understood. The SAXS-EOM method, applied here to homopolymers poly(dA) and poly(dT), is completely general and may be adapted for studying DNA or RNA

of mixed sequence. Furthermore, our methodology may be particularly useful for modeling conformations of flexible regions of RNA or DNA within a larger structural context, such as linkers between expression platform and aptamer domains in riboswitches [53] or intermediates in RNA folding [54–57].

Acknowledgments

We thank Arthur Woll and CHESS staff for experimental assistance, as well as Pollack Lab members, especially Huimin Chen, Suzette A. Pabit, Julie L. Sutton, Joshua Tokuda, and Yujie Chen for help with sample preparation, data collection, and preparation of the manuscript. Molecular models appearing in the figures were rendered using Pymol version 1.2r1 (DeLano Scientific LLC). This work was supported by the NIH (R01-GM085062). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0936384, and the MacCHESS resource is supported by the National Institute of General Medical Services (NIGMS) award GM-103485.

References

[1] M. Ullner and C. E. Woodward, "Orientational correlation function and persistence lengths of flexible polyelectrolytes", Macromolecules **35**, 1437–1445 (2002).

[2] N. M. Toan and D. Thirumalai, "On the origin of the unusual behavior in the stretching of single-stranded DNA", J. Chem. Phys. **136**, 235103 (2012).

[3] M. N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, "Stretching single stranded DNA, a model polyelectrolyte", Phys. Rev. Lett. **89**, 248102+ (2002).

[4] Y. Seol, G. M. Skinner, and K. Visscher, "Elastic properties of a single-stranded charged homopolymeric ribonucleotide", Phys. Rev. Lett. **93**, 118102+ (2004).

[5] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck, "Nonlinear low-force elasticity of single-stranded DNA molecules", Phys. Rev. Lett. **102**, 068301 (2009).

[6] D. R. Jacobson, D. B. McIntosh, and O. A. Saleh, "The snakelike chain character of unstructured RNA", Biophys. J. **105**, 2569–2576 (2013).

[7] R. P. Rambo and J. A. Tainer, "Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering", Curr. Opin. Struct. Biol. **20**, 128–137 (2010).

[8] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution", Q. Rev. Biophys. **40**, 191–285 (2007).

[9] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, "Structural characterization of flexible proteins using small-angle X-ray scattering", J. Am. Chem. Soc. **129**, 5656–5664 (2007).

[10] M. Pelikan, G. L. Hura, and M. Hammel, "Structure and flexibility within proteins as identified through small angle X-ray scattering", Gen. Physiol. Biophys. **28**, 174 (2009).

[11] M. Hammel, "Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS)", Eur. Biophys. J. **41**, 789–799 (2012).

[12] J. S. Pedersen and P. Schurtenberger, "Scattering functions of semiflexible polymers with and without excluded volume effects", Macromolecules **29**, 7602–7612 (1996).

[13] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle", Recl. Trav. Chim. Pays-Bas **68**, 1106–1122 (1949).

[14] P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. W. Ruigrok, and M. Blackledge, "A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering", Proc. Natl. Acad. Sci. U.S.A. **102**, 17002–17007 (2005).

[15] H. Eisenberg and G. Felsenfeld, "Studies of the temperature-dependent conformation and phase separation of polyriboadenylic acid solutions at neutral ph", J. Mol. Biol. **30**, 17–37 (1967).

[16] J. Witz and V. Luzzati, "La structure des acides polyadenylique et polyuridilique en solution: etude par diffusion centrale des rayons X", J. Mol. Biol. **11**, 620–630 (1965).

[17] M. A. Jonikas, R. J. Radmer, A. Laederach, R. Das, S. Pearlman, D. Herschlag, and R. B. Altman, "Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters", RNA **15**, 189–199 (2009).

[18] M. Parisien and F. Major, "The MC-fold and MC-sym pipeline infers RNA structure from sequence data", Nature **452**, 51–55 (2008).

[19] S. Yang, M. Parisien, F. Major, and B. Roux, "RNA structure determination using SAXS data", J. Phys. Chem. B **114**, 10039–10048 (2010).

[20] M. Ali, J. Lipfert, S. Seifert, D. Herschlag, and S. Doniach, "The ligand-free state of the TPP riboswitch: a partially folded RNA structure", J. Mol. Biol. **396**, 153–165 (2010).

[21] M. J. Gajda, D. M. Zapien, E. Uchikawa, and A.-C. Dock-Bregeon, "Modeling the structure of RNA molecules with small-angle X-ray scattering data", PLoS One **8**, e78007 (2013).

[22] Z. Xia, D. R. Bell, Y. Shi, and P. Ren, "RNA 3D structure prediction by using a coarse-grained model and experimental data", J. Phys. Chem. B **117**, 3135–3144 (2013).

[23] A. Gopal, Z. H. Zhou, C. M. Knobler, and W. M. Gelbart, "Visualizing large RNA molecules in solution", RNA **18**, 284–299 (2012).

[24] E. Fadrná, N. Špačková, J. Sarzynska, J. Koca, M. Orozco, T. E. Cheatham III, T. Kulinski, and J. Šponer, "Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields", J. Chem. Theory Comput. **5**, 2514–2530 (2009).

[25] S.-H. Kim, H. Berman, N. Seeman, and M. Newton, "Seven basic conformations of nucleic acid structural units", Acta Crystallogr. B **29**, 703–710 (1973).

[26] V. L. Murthy, R. Srinivasan, D. E. Draper, and G. D. Rose, "A complete conformational map for RNA", J. Mol. Biol. **291**, 313–327 (1999).

[27] L. J. Murray, W. B. Arendall, D. C. Richardson, and J. S. Richardson, "RNA backbone is rotameric", Proc. Natl. Acad. Sci. U.S.A. **100**, 13904–13909 (2003).

[28] E. Hershkovitz, E. Tannenbaum, S. B. Howerton, A. Sheth, A. Tannenbaum, and L. D. Williams, "Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA", Nucleic Acids Res. **31**, 6249–6257 (2003).

[29] G. E. Sims and S.-H. Kim, "Global mapping of nucleic acid conformational space: dinucleoside monophosphate conformations and transition pathways among conformational classes", Nucleic Acids Res. **31**, 5607–5616 (2003).

[30] B. Schneider, Z. Morávek, and H. M. Berman, "RNA conformational classes", Nucleic Acids Res. **32**, 1666–1677 (2004).

[31] M. T. Sykes and M. Levitt, "Describing RNA structure by libraries of clustered nucleotide doublets", J. Mol. Biol. **351**, 26–38 (2005).

[32] J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook, and H. M. Berman, "RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution)", RNA **14**, 465–481 (2008).

[33] E. Humphris-Narayanan and A. M. Pyle, "Discrete RNA libraries from pseudo-torsional space", J. Mol. Biol. **421**, 6–26 (2012).

[34] D. A. Erie, K. J. Breslauer, and W. K. Olson, "A Monte Carlo method for generating structures of short single-stranded DNA sequences", Biopolymers **33**, 75–105 (1993).

[35] D. Svozil, J. Kalina, M. Omelka, and B. Schneider, "DNA conformations and their sequence preferences", Nucleic Acids Res. **36**, 3690–3706 (2008).

[36] B. Schneider, S. Neidle, and H. M. Berman, "Conformations of the sugar-phosphate backbone in helical DNA crystal structures", Biopolymers **42**, 113–124 (1997).

[37] W. Saenger, *Principles of nucleic acid structure* (Springer-Verlag, 1984).

[38] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, and D. C. Richardson, "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids", Nucleic Acids Res. **35**, W375–W383 (2007).

[39] J. Applequist and V. Damle, "Thermodynamics of the one-stranded helix-coil equilibrium in polyadenylic acid", J. Am. Chem. Soc. **88**, 3895–3900 (1966).

[40] S. P. Meisburger, J. L. Sutton, H. Chen, S. A. Pabit, S. Kirmizialtin, R. Elber, and L. Pollack, "Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: beyond the wormlike chain model", Biopolymers **99**, 1032–1045 (2013).

[41] L. Pollack, "SAXS studies of ion-nucleic acid interactions", Annu. Rev. Biophys. **40**, 225–242 (2011).

[42] S. Hansen, "Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering", J. Appl. Crystallogr. **33**, 1415–1421 (2000).

[43] G. Parkinson, J. Vojtechovsky, L. Clowney, A. Brunger, and H. Berman, "New parameters for the refinement of nucleic acid-containing structures", Acta Crystallogr. D Biol. Crystallogr. **52**, 57–64 (1996).

[44] D. Cheng, L. S. Kan, E. E. Leutzinger, K. Jayaraman, P. S. Miller, and P. O. Ts'o, "Conformational study of two short pentadeoxyribonucleotides, CpCpApApG and CpTpTpGpG, and their fragments by proton NMR", Biochemistry **21**, 621–630 (1982).

[45] J. Mellema, A. Jellema, C. Haasnoot, J. Van Boom, and C. Altona, "Conformational analysis of the single-helical DNA fragment d (T-A-A-T) in aqueous solution. The combined use of NMR proton chemical shifts and coupling constants obtained at 300 MHz and 500 MHz", Eur. J. Biochem. **141**, 165–175 (1984).

[46] W. H. Press, *Numerical recipes 3rd edition: the art of scientific computing* (Cambridge university press, 2007).

[47] D. Svergun, C. Barberato, and M. H. J. Koch, "CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates", J. Appl. Crystallogr. **28**, 768–773 (1995).

[48] J. Lin, "Divergence measures based on the Shannon entropy", IEEE Trans. Inform. Theory **37**, 145–151 (1991).

[49] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA", Proc. Natl. Acad. Sci. U.S.A. **109**, 799–804 (2012).

[50] A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, "Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle X-ray scattering", Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. **86**, 021901+ (2012).

[51] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy", Biophys. J. **86**, 2530–2537 (2004).

[52] D. B. McIntosh, G. Duggan, Q. Gouil, and O. A. Saleh, "Sequence-dependent elasticity and electrostatics of single-stranded DNA: signatures of base-stacking", Biophys. J. **106**, 659–666 (2014).

[53] W. C. Winkler and R. R. Breaker, "Regulation of bacterial gene expression by riboswitches", Annu. Rev. Microbiol. **59**, 487–517 (2005).

[54] J. C. Schlatterer, L. W. Kwok, J. S. Lamb, H. Y. Park, K. Andresen, M. Brenowitz, and L. Pollack, "Hinge stiffness is a barrier to RNA folding", J. Mol. Biol. **379**, 859–870 (2008).

[55] J. H. Roh, L. Guo, J. D. Kilburn, R. M. Briber, T. Irving, and S. A. Woodson, "Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle X-ray scattering", J. Am. Chem. Soc. **132**, 10148–10154 (2010).

[56] S. A. Pabit, J. L. Sutton, H. Chen, and L. Pollack, "Role of ion valence in the submillisecond collapse and folding of a small RNA domain", Biochemistry **52**, 1539–1546 (2013).

[57] R. Russell, I. S. Millett, M. W. Tate, L. W. Kwok, B. Nakatani, S. M. Gruner, S. G. J. Mochrie, V. Pande, S. Doniach, D. Herschlag, and L. Pollack, "Rapid compaction during RNA folding", Proc. Natl. Acad. Sci. U.S.A. **99**, 4266–4271 (2002).

CHAPTER 5

MAGNESIUM IONS ALTER THE GLOBAL STRUCTURE OF SINGLE-STRANDED DNA

Steve P. Meisburger[1] Kurt Andresen[2], and Lois Pollack[1]

1. Applied and Engineering Physics, Cornell University, Ithaca, New York
2. Department of Physics, Gettysburg College, Gettysburg, Pennsylvania

Abstract

The mechanical properties of single-stranded nucleic acids play a vital role in such dynamic processes as DNA replication and RNA folding. While force spectroscopy experiments have demonstrated that duplex DNA molecules behave as ideal worm-like chains with a persistence length that can be measured accurately, the unfolded single-stranded form of DNA (ssDNA) shows diverse sequence and ion-dependent effects that defy description by simple polymer models. Using small angle x-ray scattering (SAXS), we measure the changes in global structure of ssDNA homopolymers as Mg^{2+} replaces Na^{+} in the ion atmosphere. SAXS data are interpreted using an ensemble optimization technique to visualize the conformational adaptation of the DNA backbone to changes in the ion atmosphere, including base-stacking effects. In addition, we count the number of ions associated with the ssDNA using buffer exchange atomic emission spectroscopy. For poly(dT), which shows negligible stacking interactions between bases, we observe backbone "crumpling" in the presence of Mg^{2+}. The magnitude of the orientation correlation function, reporting backbone tortuousity, decreases in proportion to the number of associated Mg ions. We also observe Mg-dependent conformational changes in poly(dA), which shows strong base-stacking preferences: Mg appears to promote long-ranged

order within B-like single-stranded helical regions. Implications for new polyelectrolyte theories of ssDNA and the biological role of electrostatics in these systems are discussed.

5.1 Introduction

The multifaceted roles that nucleic acids play in biology depend upon the structural plasticity of these molecules: the DNA double-helix melts and re-forms during transcription and replication; single-stranded RNA and DNA adopt alternate folds to transduce signals [1, 2]; and RNA is both the single-stranded messenger and the complex molecular machine that turns the genetic code into protein [3]. Double-stranded conformations are torsionally constrained and exceptionally rigid, while single-stranded forms can be highly flexible. In the single-stranded form, the sugar phosphate backbone's six torsion angles per nucleotide may adopt a variety of conformations with either stacked or unstacked bases [4, 5]. The free energy for base-stacking is sequence-dependent [6]; homopolymers with bases that favor stacking can form persistent single-stranded helices at low temperatures [7].

While rigid structures are amenable to high resolution methods such as crystallography and NMR, there are few techniques available for precisely characterizing flexible structures. Solution small angle x-ray scattering (SAXS) provides valuable information about dynamic molecules that may be integrated with computational modeling and complementary techniques [8, 9]. SAXS has been essential in understanding how RNAs fold [10, 11] and for measuring the strength of interactions between nucleic acids [12]. A persistent theme in nucleic acid research is role of electrostatics and the counterion atmosphere [13]. Mg ions fulfill an essential role in structural transitions of RNA and DNA: by interacting preferentially with negatively charged phosphate groups, Mg^{2+} compensates for electrostatic repulsion that favors unfolding. The salt-dependent equilibrium conformations of single-stranded RNA and DNA homopolymers

have been studied by small angle x-ray scattering (SAXS) [14–16], single molecule Förster resonance energy transfer spectroscopy [14, 16–18], and hydrodynamic techniques [19, 20]. In addition, the base sequence and salt environment influence ssRNA and ssDNA's mechanical properties [21–29] and dynamics [30, 31].

The conformation, dynamics, and mechanical properties of single-stranded nucleic acids are related by the underlying statistical and physico-chemical properties of the polymer. Polyelectrolyte theory for flexible charged polymers has been applied to understand the origin of salt-dependent mechanical properties and to predict equilibrium conformations. The analytically tractable charged wormlike chain theory [32–36] has been successful for describing double-stranded DNA [37], but fails to describe the mechanical flexibility of single-stranded nucleic acids in physiologically-relevant conditions [21, 26]. In particular, single molecule force extension measurements of long ssRNA and ssDNA show a high-force scaling law of $x \sim \log(f)$ that is inconsistent with wormlike chain theory [21, 26]. Computational studies using mean-field electrostatics were combined with scaling arguments to show that this $x \sim \log(f)$ behavior of flexible polyelectrolytes is related to a power-law decay of the tangent vector orientation correlation function [38]. In the wormlike chain theory, orientation correlation functions decay exponentially: thus, a power-law decay orientation correlation function may be regarded as the defining feature for a distinct class of highly flexible polyelectrolytes, termed "snakelike" chains [26]. In Chapter 4, we used SAXS with a novel ensemble optimization method (EOM) to confirm that the ssDNA homopolymer poly(dT) indeed shows the signature power-law decay of the orientation correlation function in 20 mM NaCl solution.

The physical origin of the power-law decay (and associated $x \sim \log(f)$ scaling) is still actively debated. Simulations of ssDNA-like polyelectrolytes with explicit ions attributed the logarithmic high-force scaling to ion-associated clusters [39]. On the other hand, mean field simulations (without explicit ions) suggest that the power-law dependence arises from the sub-

stantial overlap of the non-electrostatic intrinsic persistence length and the length-scales associated with electrostatic or excluded volume effects [38]. Thus, discrete ion effects are not necessary to explain the conformations of single-stranded DNA in monovalent salts. In contrast, force-extension curves of single-stranded nucleic acids in divalent salts (Mg^{2+} and Ca^{2+}) show extra compliance at high forces [27], and mean-field theories that include electrostatics only through ionic strength cannot account for this ion valence effect. Thus, it has been suggested that divalent ions interact strongly and specifically with phosphate groups. General models have been proposed, including a picture where the chain "wraps" around counterions [27], and another where it adopts a "crumpled" structure in the presence of these ions [39]. These models have yet to be confirmed by structural measurements.

Recently, we measured the electrostatic effect of divalent ions on poly(dT) [16]. The repulsion between chains is dramatically reduced when $MgCl_2$ is present. In the context of approximate mean-field electrostatic theory, this was explained by a charge-renormalizing effect due to condensation of Mg^{2+} in a region of high potential ($\Phi > k_B T/e$) near the molecule. A freely-rotating chain model with screened coulomb interactions was sufficient to describe the overall size of ssDNA coils measured using SAXS and smFRET. High-q features of the SAXS curves, sensitive to the internal structure of the ssDNA coils, resembled FRC model predictions (here, $q = 4\pi \sin(2\theta/2)/\lambda$ where 2θ is the scattering angle and λ is the wavelength). However, agreement was not exact. Several studies have attempted to fit models directly to SAXS data in order to learn about the internal structure of polymers. For flexible proteins, the fractal dimension has been found using a power-law fit at mid-q [40, 41]; however, this fitting technique is not applicable to ssDNA because critical length scales overlap and are non-separable in q-space: the persistence length, phosphate-phosphate distance, and chain diameter are typically of order 10 Å. SAXS data for single-stranded RNA and DNA in monovalent salts have also been fit using a semi-analytic wormlike chain form factor [14, 15, 42]. The wormlike chain model neglects long-ranged interactions and is only appropriate for ssDNA at high ionic strengths where elec-

trostatic interactions are completely screened [21]. Furthermore, the assumption of isotropic flexibility implicit in polymer theories is unlikely to be true for ssDNA when base-stacking is present.

To provide a general method for extracting molecular details from SAXS data for ssDNA, we developed an efficient atomistic modeling routine to generate ssDNA structure pools and select ensembles that fit the SAXS data (Chapter 4). From the fitted ensembles, we obtain the orientation correlation function of the chain, which contains the essential information about the internal structure of the polymer (i.e. static flexibility) for comparison with theory. This SAXS-based ensemble optimization method (SAXS-EOM) was also applied to poly(dA), whose adenine bases have favorable stacking interactions, demonstrating the generality of the technique.

Here, we seek to understand the effects of Mg^{2+} on the solution structure of ssDNA. We interpret x-ray scattering measurements of poly(dA) and poly(dT) using SAXS-EOM to obtain the orientation correlation function. In addition, the affinity of Mg^{2+} for ssDNA relative to Na^{+} is investigated by ion counting using buffer exchange atomic emission spectroscopy (BE-AES) [43]. We examine changes in the SAXS-EOM ensembles that correlate with the number of associated Mg ions. We find that the effects of Mg^{2+} are sequence-dependent. For poly(dT), we find that the magnitude of the correlation function (related to backbone tortuosity) depends linearly on the number of Mg ions. For poly(dA), Mg^{2+} promotes long-ranged order in base-stacked single-stranded helices. Remarkably, the distinct structural effects of Mg^{2+} on poly(dT) and poly(dA) do not manifest in different Mg^{2+} affinities: the number of Mg^{2+} ions associated with poly(dT) and poly(dA) is identical within experimental uncertainty in all solution conditions measured. Our measurements provide a structural basis for the increased compliance of unstacked chains in the presence of divalent ions [27] and suggest future experiments to test predictions of the EOM models that remain speculative because of the low resolution of the SAXS measurement.

5.2 Materials and Methods

5.2.1 Sample preparation

Four DNA oligonucleotides were synthesized and HPLC purified by Integrated DNA Technologies (Coralville, IA); sequence GCATCTGGGCTATAAAGGGCGTCG (S1), its complement (S2), a 30-mer of deoxythymidine (dT30); and a 30-mer of deoxyadenosine (dA30). Lyophilized strands S1 and S2 were re-suspended in STE buffer (10 mM TRIS, 50 mM NaCl, 1 mM EDTA, pH 8.0) at a concentration of 0.5 mM. To generate a 25 base-pair DNA duplex (ds25), strands S1 and S2 were mixed together in an equimolar ratio in a microcentrifuge tube, placed in a 95C water bath for 5 minutes, and allowed to cool on the bench. Single-stranded DNA homopolymers were re-suspended in TE buffer (10 mM TRIS, 1 mM EDTA, pH 8.0). For SAXS experiments, DNA was exchanged five times using Amicon Ultra 0.5, 3kDa-cutoff centrifugal concentrators (EMD Millipore, Billerica, MA) at $10,000 \times g$ with buffers containing 1 mM Na+MOPS, 20 mM NaCl, and 0, 1, 2, 5, 10, or 20 mM $MgCl_2$. The matching buffer was retained for background subtraction and dilution at the beamline.

For ion-counting experiments, samples were equilibrated using centrifugal concentrators with buffers containing 10 mM Na+MOPS pH 7 and added NaCl ($[Na^+]= 20\,mM$) and either 0, 0.5, 1, 2, 5, 10, 15, or 20 mM $MgCl_2$. Eight dilution and concentration cycles were carried out with the volume of the concentrate kept above $100\,\mu L$ [44]. After the final concentration cycle, the concentrate and flow-through were immediately diluted 500-750 fold to concentrations appropriate for the ICP-AES instrument: for each dilution 20-30 μL was added to 15 mL of 10 mM high-purity ammonium acetate (Sigma-Aldrich) to promote solubility of trace metal ions [43]. For all samples with bulk Mg concentrations between 0 and 10 mM, the DNA concentration pre-dilution (determined by ICP-AES) was 0.39-0.46 mM for dT30, 0.45-0.53 mM for dA30, and 0.16-0.32 mM for ds25. Higher DNA concentrations were used for the 15

mM and 20 mM Mg buffer conditions to improve the signal-to-background ratio: 1.1-1.2 mM for dT30, 1.5-1.9 mM for dA30, and 0.78-1.2 mM for ds25. Sources of uncertainty in the ion-counting experiment include pipetting error in the dilution step and detection noise or instability of the ICP-AES instrument. Therefore, 2-4 separate ICP data sets with independent calibrations were acquired; several atomic emission lines were recorded for each element (3 for P, 2 for Na, and 4 for Mg); and 3-4 separate dilutions were performed when feasible (excluding the DNA-containing samples with 15 and 20 mM Mg, where sample quantity limited the number of dilutions to one).

5.2.2 Ion-counting with buffer exchange atomic emission spectroscopy

Concentrations of the counterions and DNA were determined using an Optima 7300DV ICP-AES (Perkin Elmer, Waltham, MA) within the linear detection range of the instrument: emission lines for elements P, Na, and Mg were monitored (Cl^- anions were not detected) and integrated intensities were converted to concentration units by calibrating the instrument with standard solutions. The number of excess ions per phosphate was determined for each sample using

$$\Delta N_{Ion}/N_P = \frac{[\mathrm{Ion}]_S - [\mathrm{Ion}]_B}{[\mathrm{P}]_S - [\mathrm{P}]_B} \tag{5.1}$$

where the subscripts S and B, refer to the DNA-containing sample and the corresponding buffer flow-through, respectively. The competition curves were fit using a 4-parameter phenomenological model from Ref. [44] (Hill-equation):

$$\Delta N_{Na}/N_P = \frac{F_{Na}}{1 + ([\mathrm{Mg}]/M_{1/2})^n} \tag{5.2}$$

$$\Delta N_{Mg}/N_P = \frac{F_{Mg}([\mathrm{Mg}]/M_{1/2})^n}{1+([\mathrm{Mg}]/M_{1/2})^n} \tag{5.3}$$

where [Mg] is the bulk Mg concentration, $M_{1/2}$ is the competition constant, n is the Hill coefficient, F_{Na} is the excess Na/P ratio at $[\mathrm{Mg}] = 0$, and F_{Mg} is the Mg/P ratio in the limit $[\mathrm{Mg}] \to \infty$. When analyzing the data, each ICP-AES reading was considered an independent measurement. Equations 5.2 and 5.3 were simultaneously fit to the data using the lsqnonlin function in Matlab (MathWorks, Natick, MA), where the points were weighted according to the standard deviation of all measurements for that sample, and errors in the fit parameters were estimated by bootstrapping [45]. For presentation purposes, data points for each sample were averaged, and the error bars were computed as the standard deviation divided by the square root of the number of measurements.

5.2.3 X-ray scattering measurements and ensemble optimization

SAXS data collection was described in Chapter 4 (see Section 4.2.1) for poly(dT) and poly(dA) in Na-only solutions. Here, data for the Mg-containing solutions were acquired identically. Briefly, interparticle interference effects were corrected by extrapolation to the dilute limit using SAXS profiles for a series of DNA concentrations: 50, 100 and 200 μM. Ensemble optimization using an all-atom rotamer model for DNA, developed in Chapter 4, was employed. Torsion angle weights were re-optimized for each solution condition. Final pools of 10,000 structures were generated using the optimized weights, and 500 ensembles of 20 structures that fit each SAXS profile were selected by the genetic algorithm method GAJOE13 [8, 46]. Orientation correlation functions were computed for virtual bonds between phosphorus atoms, also described in Chapter 4 (see Section 4.3).

5.2.4 Theoretical prediction of ion numbers from Poisson-Boltzmann calculations

Nonlinear Poisson-Boltzmann (NLPB) calculations were performed for duplex DNA, dA_{30}, and dT_{30}. An ideal B-form duplex model was generated using Nucleic Acid Builder [47]. For single-stranded DNAs, the best-fit ensemble of 20 structures for each molecule was used. Hydrogen atoms, Van der Waals radii, and partial charges were assigned using the AMBER parameter set by PDB2PQR [48]. For electrostatics calculations, the program APBS [49] was used with a Stern layer width (ion radius) of 2 Å, a dielectric constant of 2 inside and 78.54 outside the molecule, solvent probe radius of 1.4 Å, a grid spacing of < 1 Å, and zero-potential boundaries placed 40 Å from the molecular surface. After solving the NLPB equation, the number of excess ions of each type within the calculation box, N_i^{in}, was calculated from the reduced potential ϕ_j (units k_BT/e) at grid points $j \in \Pi$, where Π is the list of ion-accessible sites;

$$N_i^{in} = A_0\, c_i^{\infty}\, V_{cell} \sum_{j\in\Pi} \left(\exp(-z_i\phi_j) - 1\right)$$

where $A_0 = 6.022 \times 10^{-4} M^{-1} \text{Å}^{-3}$ is a unit-conversion constant for bulk ion concentration c_i^{∞} in molar, and cell volume V_{cell} in Å^3. With excess ion numbers computed in this way, the simulation box may not be sufficiently large to include all excess ions. Thus, ion numbers were corrected by applying the linear Poisson Boltzmann (LPB) equation in the low-potential region "outside the box" together with the charge-neutrality condition $Z_{DNA} + \sum_i z_i N_i = 0$. In LPB, each ion neutralizes charge in proportion to its contribution to the ionic strength of the solution [50]. Thus, corrected ion numbers N_i that are robust to the choice of boundary conditions were calculated using,

$$N_i = N_i^{in} + \left(Z_{DNA} - \sum_j N_j^{in} z_j\right) \frac{z_i c_i}{\sum_j z_j^2 c_j}$$

5.3 Results and Discussion

In a mixed-salt environment, counterions compete for association with DNA. This competition has been studied using equilibrium dialysis experiments where the bulk concentration of one counterion is fixed while the other is varied. Previous dialysis experiments measuring ion competition around double-stranded DNA using BE-AES chose a standard condition of 20 mM NaCl with varying $MgCl_2$ [44]. The low monovalent ion concentration is important for measuring the Na concentration difference with sufficient signal-to-noise. Therefore, we adopted the same standard monovalent ion condition. First, we repeated these experiments using a synthetic 25 base-pair DNA duplex (charge of -48e) that has been the subject of previous x-ray scattering studies (e.g. Refs. [12, 51–53]). The competition curves are reported in Figure 5.1c, and parameters that were fit using Equations 5.2 and 5.3 are given in Table 5.1. For double-stranded DNA, we find a competition constant of $M_{1/2} = 0.48 \pm 0.01$ mM that reproduces earlier BE-AES measurements for DNA duplexes with a different base sequence but comparable length [44]. Next, we measured the competitive association of Mg^{2+} to two single-stranded homopolymers dT_{30} (-29e), and dA_{30} (-29e). As was the case with double-stranded DNA, Mg^{2+} readily competes with 20 mM Na^+ (Figures 5.1a-b). The numbers of excess Mg ions associated with dT_{30} and dA_{30} at each bulk $MgCl_2$ concentration are identical within experimental uncertainty, but distinct from the numbers associated with double-stranded DNA (Figure 5.1d): the concentration of $MgCl_2$ required to out-compete half of the Na^+ ions associated with single-stranded DNA is approximately two-fold greater than is required to out-compete the same number of Na^+ ions around double-stranded DNA ($M_{1/2}$, Table 5.1).

Single-stranded DNA's flexibility causes salt-dependent swelling for chains in the size-range studied here (c.f. Ref. [15]). To correlate the competition between Na^+ and Mg^{2+} with changes in ssDNA conformation, we collected SAXS data for dT_{30} and dA_{30} with solution conditions nearly identical to the BE-AES experiments. SAXS profiles were extrapolated to infinite dilu-

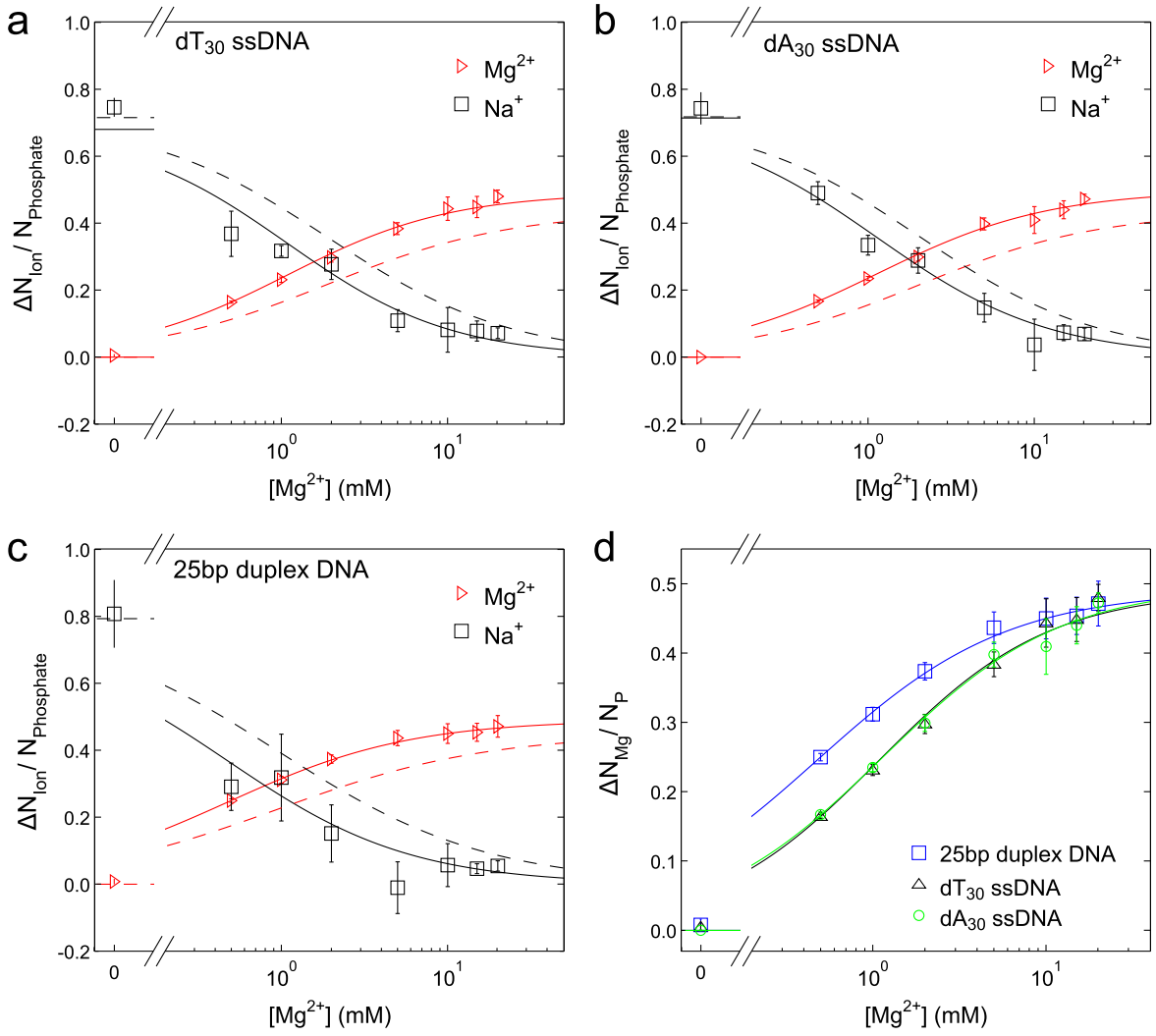

Figure 5.1: BE-AES measurements of counterion competition around single- and double-stranded DNA. The number of excess Mg and Na ions per DNA phosphate were measured with a fixed Na bulk concentration of 20 mM and varying $MgCl_2$ concentration between 0 and 20 mM. BE-AES data are shown for dT_{30} (a), dA_{30} (b), and 25bp duplex DNA (c). Solid lines show the best fit of equations 5.2 and 5.3 to the data (parameters in Table 5.1), and dashed lines are the nonlinear Poisson Boltzmann calculations. The number of excess Mg ions for the three DNA molecules in (a-c) are re-plotted in (d) for visual comparison.

	$M_{1/2}$(mM)	n	F_{Mg}	F_{Na}
ds25	0.48 ± 0.01	0.79 ± 0.02	0.491 ± 0.004	0.74 ± 0.03
dt30	1.09 ± 0.04	0.88 ± 0.02	0.490 ± 0.007	0.68 ± 0.02
da30	1.13 ± 0.03	0.84 ± 0.01	0.498 ± 0.005	0.71 ± 0.01

Table 5.1: Fit parameters for Mg-Na competition around single- and double-stranded and DNA (Equations 5.2 and 5.3)

tion to account for interparticle interference effects as described in Methods, and are plotted in Figures 5.2a (dT_{30}) and 5.2c (dA_{30}).

Using an ensemble optimization method specifically developed for poly(dT) and poly(dA), we fit the SAXS data with ensembles of atomically-detailed models. Although the ensembles contain high resolution information, it is important to interpret these ensembles on length-scales that are constrained by the limited resolution of the X-ray data. We previously used orientation correlation functions to interpret the flexibility of these homopolymers in 20 mM NaCl with no Mg^{2+} (Chapter 4). Here, orientation correlation functions were additionally found from SAXS data with Mg^{2+} (Figure 5.2).

For dT_{30}, we find that the orientation correlation functions have a characteristic power-law decay over a range of about 10 nucleotides, given by [38]

$$\left\langle \cos\theta_{i,j} \right\rangle = C\,|i-j|^{-\gamma} \tag{5.4}$$

The best-fit values for the exponent γ are similar for all Mg concentrations, while the magnitude C decreases with added Mg. Since C is associated with the backbone curvature, and γ with the excluded-volume interaction [38], the correlation functions for poly(dT) suggest that the backbone becomes more "crumpled" when Mg^{2+} counterions replace Na^{+}. Increased backbone tortuosity is evident when visually comparing the atomic models for dT_{30} in Mg-free and Mg-saturated conditions (Figure 5.3b and Supplemental Figure B.1). In the models, crumpled conformations arise from highly-deformed torsion angle ranges such as a1p, where torsion angles ξ and α are in the non-canonical *gauche+* (see Table 4.1 in Chapter 4 and Supplemental Figure B.2b).

The orientation correlation functions for dA_{30}, shown in Figure 5.2c, have an oscillatory behavior that is not well-described by Equation 5.4. The correlation function oscillations are due to the selection of models with stacked conformations that twist the backbone in a right-handed helix, exemplified by the structures in Figure 5.3d. This helical oscillation is superimposed on

the long-ranged orientation correlation of segments due to electrostatic and excluded volume effects. The depth of the first negative-going peak around $|i-j| \sim 4$ deepens monotonically with $MgCl_2$ concentration, suggesting that Mg^{2+} promotes well-ordered and persistent curvature of the backbone. A closer examination of the torsion angle frequencies in the optimized dA_{30} pools reveals that the frequency of the BI-type base stack doubles between 0 and 20 mM $MgCl_2$ (Supplemental Figure B.2a). At the highest Mg concentrations, $\sim 80\%$ of backbone torsion angles in the ensembles have the BI conformation (note that the overall number of stacked residues is not strongly Mg-dependent, as other stacked conformations such as BII are present in 0 mM $MgCl_2$: see Supplemental Figure B.2a). The positive-going peak around $|i-j| \sim 8$ first increases showing the effects of persistent helices, and then decreases slightly for [$MgCl_2$]$>$ 5 mM as the chain assumes more compact dimensions in the well-screened salt environment.

By combining the SAXS and BE-AES data, we gain additional insight into the specific role of Mg^{2+} in promoting conformational changes in single-stranded DNA. The orientation correlation function parameters C and γ for dT_{30} are plotted as a function of the number of associated Mg^{2+} ions per phosphate group in 5.3a. We find a strong linear correlation between C and the number of excess Mg^{2+} measured by BE-AES that suggests a specific role for Mg^{2+} in stabilizing the "crumpled" conformations. In contrast, ssDNA-like polyelectrolyte simulations using mean-field electrostatics found that the magnitude C in Equation 5.4 was independent of ionic strength [38]. Thus, our observation of an ion-dependent C (and constant γ) provides further evidence that specific interactions with Mg^{2+}, as opposed to diffuse screening effects, are important.

Similar insight into the specific role of Mg^{2+} in altering the structure of dA_{30} are found from combining BE-AES data with the SAXS ensembles. In the ensembles, base-stacked helices in NaCl lack long-ranged order because of a diversity of stacked conformations. These helices become more ordered at high $MgCl_2$ concentrations where a single stacked conformer (BI)

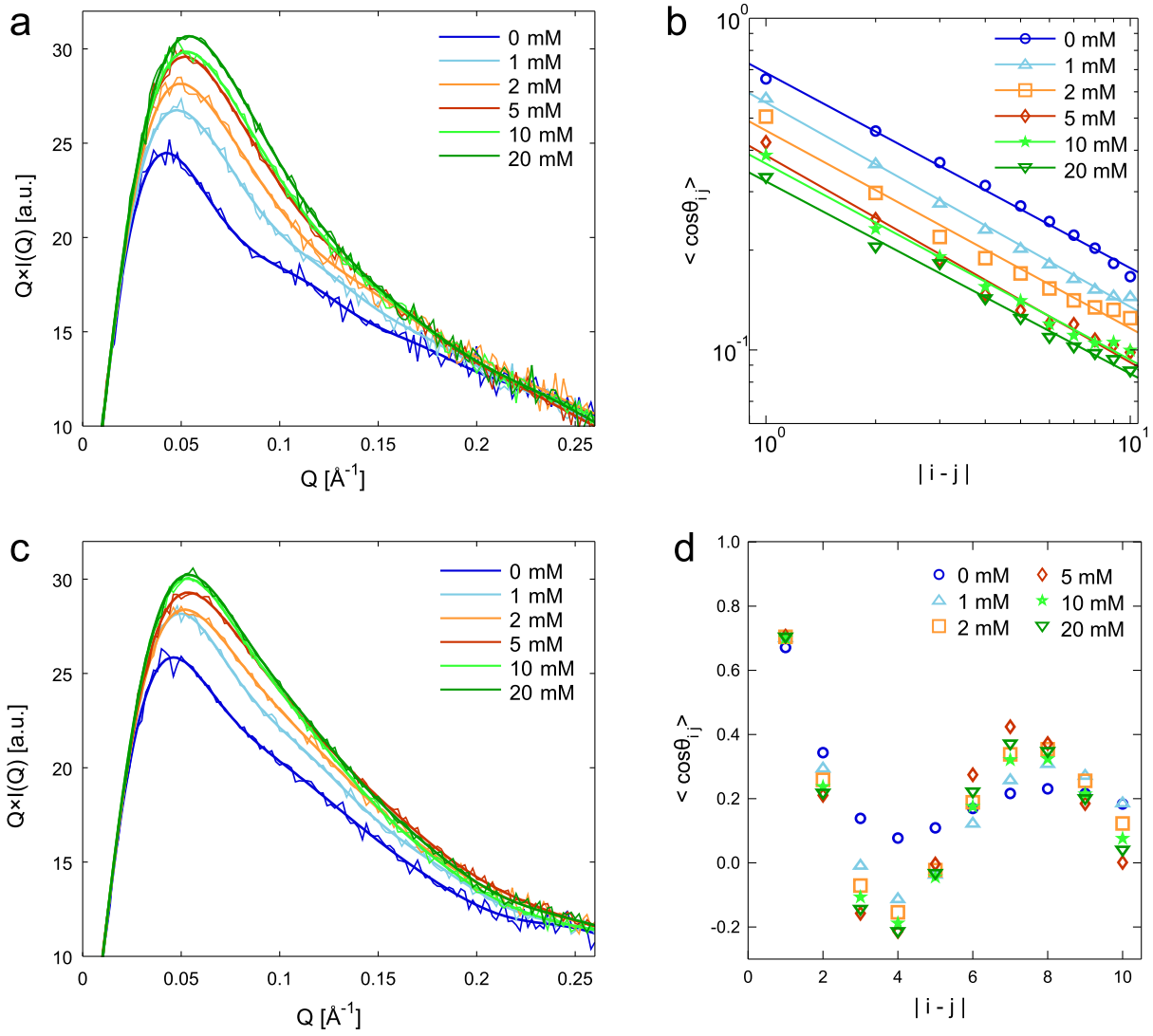

Figure 5.2: Mg-dependent orientation correlation function for single-stranded DNA. (a) X-ray scattering data for dT_{30} with 20 mM NaCl and increasing $MgCl_2$ indicated in the legend is plotted as $q \times I(q)$ vs. q to emphasize the high-angle region ($q = 4\pi \sin(2\theta/2)/\lambda$ with wavelength λ and scattering angle 2θ). Fits using ensemble optimization are superimposed. (b) Orientation correlation function analysis of the dT_{30} ensembles. Virtual bonds between phosphate atoms defined the backbone tangent vectors. The average dot-product of normalized tangent vectors i and j is $\langle \cos\theta_{ij} \rangle$, were the brackets signify the mean for constant separation $|i-j|$. $MgCl_2$ concentrations are indicated in the legend. Correlation functions were fit with a power-law decay in the region $1 \leq |i-j| \leq 10$ (solid lines). For details, see the text. (c) Same as (a), but with dA_{30} data. (d) Orientation correlation functions for dA_{30} as in (b). Solid lines interpolate data points to guide the eye.

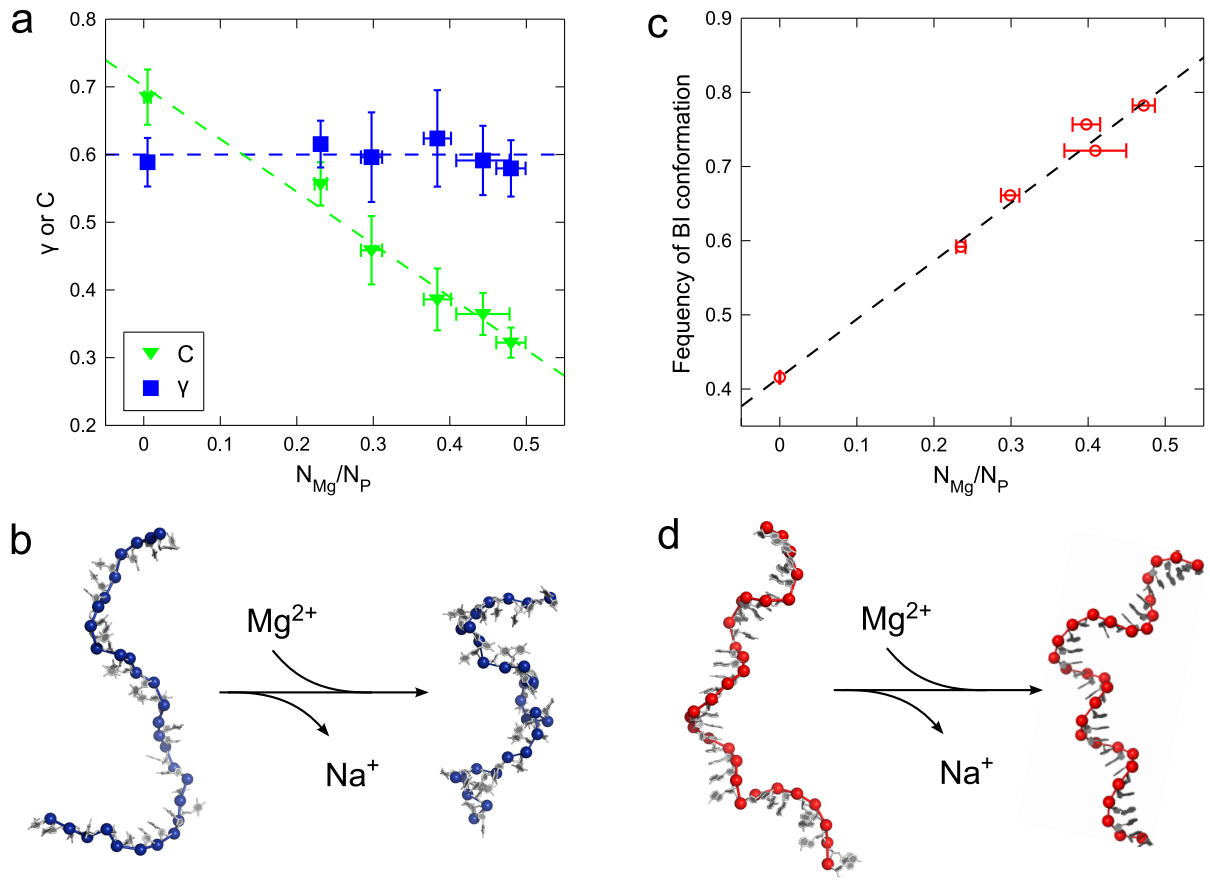

Figure 5.3: Sequence-dependent effects of Mg^{2+} on the global structure of ssDNA. (a) Power-law exponent γ and magnitude C of the orientation correlation function for dT_{30} (Figure 5.2b) are plotted as a function of the number of excess Mg^{2+} ions per DNA phosphate, N_{Mg}/N_P, determined from BE-AES (Figure 5.1a). Dashed lines show linear fits: $\gamma \sim 0.6$ and $C \sim 0.7 - 0.78\,N_{Mg}/N_P$. (b) Example dT_{30} conformations from the best-fit ensembles in NaCl only (left) and with 10 mM $MgCl_2$ added (right) with phosphorus atoms drawn as blue spheres rendered using Pymol version 1.2r1 (DeLano Scientific LLC). For the full ensembles, see Figure B.1. (c) Exchange of Na^+ for Mg^{2+} is associated with increased purity of base-stacked conformations in the ensembles. The prevalence of the canonical *BI*-type stack doubles when $MgCl_2$ is the dominant counterion (see also Figure B.2a). The dashed line shows a linear fit $f_{BI} \sim 0.78\,(N_{Mg}/N_P - 0.53)$. (d) Example dA_{30} conformations are shown from the NaCl only (left) and 20 mM $MgCl_2$ (right) ensembles.

dominates. When the frequency of the BI backbone conformation is plotted as a function of the number of associated Mg^{2+} ions per phosphate group, a linear relationship is again observed (Figure 5.3c), suggesting that Mg^{2+} directly promotes long-ranged order in single-stranded helices.

Despite differences in the conformations of poly(dT) and poly(dA), the relative affinity of Mg^{2+} vs. Na^{+} is found to be identical in both cases (Figure 5.1d), suggesting that the distinct conformational changes observed for poly(dT) and poly(dA) in Mg-containing solutions do not contribute significantly to the free energy of Mg association. Therefore, we performed nonlinear Poisson-Boltzmann (NLPB) calculations using the single-stranded DNA models from the best-fit ensemble at each solution condition in order to predict the competition curves. The results are shown as dashed lines in Figure 5.1a-c. We find that the competition constant of Mg^{2+} measured by BE-AES is more than a factor of two less than predicted by NLPB. Previous measurements of counterion competition around duplex DNA [44], successfully repeated here, also disagree with NLPB by a similar factor (Figure 5.1c). It has been argued that NLPB fails for Mg/Na competition because of neglected ion-ion correlations [54]. If that were the case, NLPB should provide an improved prediction for ssDNA, which has a low linear charge density, relative to double-stranded DNA. However, our measurements show that this is not the case. Instead, it seems that the failure of NLPB for Mg/Na competition is unrelated to counterion density in the condensed region. Rather, it is likely that the NLPB does not adequately account for the electrostatic environment of the ions that are close to the molecular surface. A similar conclusion was found in a computational study of ion distributions around an RNA duplex that correlated molecular dynamics and NLPB calculations: satisfactory agreement was found only when the NLPB calculation was modified phenomenologically with a distance-dependent dielectric constant [55].

5.4 Conclusions and Future Work

Mg ions alter the global structure of single-stranded DNA. SAXS-EOM ensembles show that poly(dT) has a crumpled structure in solutions with Mg^{2+} that is distinct from the structure in NaCl. The tortuous chains observed in SAXS-EOM ensembles in the presence of Mg^{2+} are reminiscent of the "crumpled" structures of polyelectrolytes seen in explicit ion simulations that recapitulate force extension curves [39]. The extra compliance of poly(dT) in Mg^{2+} observed by single-molecule force spectroscopy [27] is consistent with a force-dependent transition from the crumpled backbone conformers stabilized by Mg^{2+} to the more extended conformers preferred in Mg^{2+}-free solutions.

Using BE-AES, we find that $\sim$ 1 mM Mg^{2+} is sufficient to out-compete half of the Na^{+} from around single-stranded DNA when Na^{+} is present at a bulk concentration of 20 mM. The BE-AES data support our previous observation of an apparent Mg^{2+} competition constant of $\sim$ 1 mM when the second virial coefficient of poly(dT) in mixed-salt solutions was analyzed using a simple polymer model with a charge-renormalized, screened Coulomb potential [16]. However, the BE-AES data do not agree with theoretical predictions from NLPB. We note that a similar factor-of-two disagreement is also seen for duplex DNA structures here and in previous studies [44].

The competition constant for Mg^{2+} relative to Na^{+} for ssDNA does not depend on the prevalence of base-stacked residues. SAXS-EOM models of dA_{30} show a partially stacked but flexible conformation, consistent with force spectroscopy measurements of poly(dA) that show low-force flexibility and a high-force structural transition attributed to the unstacking of bases [29]. Unlike dA_{30}, dT_{30} shows a highly variable structure with mostly unstacked bases in SAXS-EOM ensembles: consistent with this observation, ssDNA lacks an unstacking transition at high forces [21, 27].

Thus, we observe two DNA sequence-dependent effects that are seemingly incompatible. On the one hand, poly(dT) and poly(dA) show distinct conformational responses to the presence of Mg^{2+} in their ion atmospheres. On the other, competition curves measured by BE-AES for dA_{30} and dT_{30} are identical within experimental uncertainty. This raises the important question of how Mg^{2+} interacts with single-stranded DNA of differing sequence.

The crumpled structure of poly(dT) with saturating Mg^{2+} is suggestive of ion-bridging interactions between phosphate groups. Hydration studies by Kankia show that the exchange of Na^{+} counterions for Mg^{2+} around poly(dT) is accompanied by the net release of water, which was interpreted as a partial dehydration of Mg^{2+} by coordination between a pair of phosphates [56]. Poly(dA) does not form a crumpled structure, presumably because base-stacking locks the backbone in a single conformation range. Furthermore, we see no evidence for long-ranged Mg-bridging interactions in the SAXS-EOM ensembles. Thus, poly(dA) should be unable to adapt its structure in order to coordinate the Mg ion. In accord with this observation, hydration measurements find no net release of water for Mg^{2+} interactions with poly(dA) [56].

The ion-bridging hypothesis would predict vastly different Mg^{2+} affinities for poly(dT) and poly(dA), however we find experimentally that they are identical. It is difficult to imagine a compensatory mechanism that would allow poly(dA) to have an identical affinity for Mg^{2+} if poly(dT) binds divalent ions directly and poly(dA) does not. Thus, we are forced to reconsider the coordination hypothesis, despite its intuitive appeal.

We propose that Mg^{2+} associates diffusively (i.e. as a fully-hydrated ion) with both poly(dT) and poly(dA). We speculate that the observation by Kankia of water release accompanying the Mg^{2+}-poly(dT) interaction is caused by changes in overall *molecular* hydration due to the crumpled backbone, rather than changes in ion hydration. The connection between hydration and the backbone conformations of DNA are well documented for duplex structures (e.g. A, B, and Z-forms show hydration-dependent stabilities [6]). Furthermore, when Mg^{2+} is observed

in crystal structures of DNA, it is fully hydrated (c.f. PDB 1D23).

If poly(dT) interacts diffusively with Mg^{2+}, why does the chain crumple? One possibility is that the crumpled structure represents the preferred backbone conformation ensemble when electrostatic interactions are well-screened. However, the non-equivalence of mono- and di-valent salts in force extension measurements [27] tends to discount this hypothesis. It is possible that backbone crumpling of poly(dT) is a hydration effect that is linked with Mg^{2+} association, as seen for example in RNA folding [57]. The structural effects of hydration on unfolded proteins have been investigated using osmolytes to alter the water activity (c.f. Ref [58]). Similar experiments have been carried out to a limited extent with ssDNA. Intriguingly, force extension curves of ssDNA in urea suggest that disrupting hydration increases the apparent length of the chain [28]. We suggest that the recently-articulated formalism for deriving ion uptake from the analysis of force extension data sets in varying salt concentrations [59, 60] might be applied to the problem of water uptake by varying the osmolyte concentration, with proper attention to the electrostatic properties of the osmolyte [61] and ion-water linkage [57].

In summary, we applied an atomistic modeling approach to interpret x-ray scattering data from the single-stranded DNA molecules dT_{30} and dA_{30} in salt solutions with varying $MgCl_2$ and we counted the number ions associated with the DNA in each condition. We provide the first evidence from a structural technique that ssDNA has a "crumpled" conformation in the presence of Mg^{2+}. Future studies will extend this work to ssRNA, previously shown to have different flexibility and base-stacking properties from ssDNA [14, 26], in order to understand the Mg-dependent flexibility of single-stranded linkers in the context of RNA folding [62, 63].

Acknowledgments

We thank members of the Pollack Lab for help with x-ray sample preparation and data collection, especially Peter Gu for preparing BE-AES samples, CHESS staff and beamline scientists Arthur Woll and Richard Gillilan for experimental assistance, and David Case and Hung Nguyen for insightful discussions. This work was supported by the NIH (R01-GM085062). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0936384, and the MacCHESS resource is supported by the National Institute of General Medical Services (NIGMS) award GM-103485.

References

[1] R. V. Brown and L. H. Hurley, "DNA acting like RNA", Biochem. Soc. Trans. **39**, 635 (2011).

[2] A. Serganov, "The long and the short of riboswitches", Curr. Opin. Struct. Biol. **19**, 251–259 (2009).

[3] T. R. Cech, "The ribosome is a ribozyme", Science **289**, 878–879 (2000).

[4] J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook, and H. M. Berman, "RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution)", RNA **14**, 465–481 (2008).

[5] D. Svozil, J. Kalina, M. Omelka, and B. Schneider, "DNA conformations and their sequence preferences", Nucleic Acids Res. **36**, 3690–3706 (2008).

[6] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, *Nucleic acids: structures, properties, and functions* (University science books, 2000).

[7] H. Eisenberg and G. Felsenfeld, "Studies of the temperature-dependent conformation and phase separation of polyriboadenylic acid solutions at neutral ph", J. Mol. Biol. **30**, 17–37 (1967).

[8] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun, "Structural characterization of flexible proteins using small-angle X-ray scattering", J. Am. Chem. Soc. **129**, 5656–5664 (2007).

[9] R. P. Rambo and J. A. Tainer, "Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering", Curr. Opin. Struct. Biol. **20**, 128–137 (2010).

[10] J. H. Roh, L. Guo, J. D. Kilburn, R. M. Briber, T. Irving, and S. A. Woodson, "Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle X-ray scattering", J. Am. Chem. Soc. **132**, 10148–10154 (2010).

[11] R. Russell, I. S. Millett, M. W. Tate, L. W. Kwok, B. Nakatani, S. M. Gruner, S. G. J. Mochrie, V. Pande, S. Doniach, D. Herschlag, and L. Pollack, "Rapid compaction during RNA folding", Proc. Natl. Acad. Sci. U.S.A. **99**, 4266–4271 (2002).

[12] X. Qiu, L. W. Kwok, H. Y. Park, J. S. Lamb, K. Andresen, and L. Pollack, "Measuring Inter-DNA potentials in solution", Phys. Rev. Lett. **96**, 138101+ (2006).

[13] V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, "A repulsive field: advances in the electrostatics of the ion atmosphere", Curr. Opin. Chem. Biol. **12**, 619–625 (2008).

[14] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded RNA and DNA", Proc. Natl. Acad. Sci. U.S.A. **109**, 799–804 (2012).

[15] A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, "Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle X-ray scattering", Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. **86**, 021901+ (2012).

[16] S. P. Meisburger, J. L. Sutton, H. Chen, S. A. Pabit, S. Kirmizialtin, R. Elber, and L. Pollack, "Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: beyond the wormlike chain model", Biopolymers **99**, 1032–1045 (2013).

[17] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, "Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy", Biophys. J. **86**, 2530–2537 (2004).

[18] T. A. Laurence, X. Kong, M. Jäger, and S. Weiss, "Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins", Proc. Natl. Acad. Sci. U.S.A. **102**, 17348–17353 (2005).

[19] S. Doose, H. Barsch, and M. Sauer, "Polymer properties of polythymine as revealed by translational diffusion", Biophys. J. **93**, 1224–1234 (2007).

[20] J. B. Mills, E. Vacano, and P. J. Hagerman, "Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA)", J. Mol. Biol. **285**, 245–257 (1999).

[21] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck, "Nonlinear low-force elasticity of single-stranded DNA molecules", Phys. Rev. Lett. **102**, 068301 (2009).

[22] Y. Seol, G. M. Skinner, and K. Visscher, "Elastic properties of a single-stranded charged homopolymeric ribonucleotide", Phys. Rev. Lett. **93**, 118102+ (2004).

[23] Y. Zhang, H. Zhou, and Z. C. Ou-Yang, "Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions.", Biophys. J. **81**, 1133–1143 (2001).

[24] M. N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette, "Stretching single stranded DNA, a model polyelectrolyte", Phys. Rev. Lett. **89**, 248102+ (2002).

[25] C. Ke, M. Humeniuk, H. S-Gracz, and P. E. Marszalek, "Direct measurements of base stacking interactions in DNA by single-molecule atomic-force spectroscopy", Phys. Rev. Lett. **99**, 018302 (2007).

[26] D. R. Jacobson, D. B. McIntosh, and O. A. Saleh, "The snakelike chain character of unstructured RNA", Biophys. J. **105**, 2569–2576 (2013).

[27] D. McIntosh and O. Saleh, "Salt species-dependent electrostatic effects on ssDNA elasticity", Macromolecules **44**, 2328–2333 (2011).

[28] S. Cui, C. Albrecht, F. Kühner, and H. E. Gaub, "Weakly bound water molecules shorten single-stranded DNA", J. Am. Chem. Soc. **128**, 6636–6639 (2006).

[29] D. B. McIntosh, G. Duggan, Q. Gouil, and O. A. Saleh, "Sequence-dependent elasticity and electrostatics of single-stranded DNA: signatures of base-stacking", Biophys. J. **106**, 659–666 (2014).

[30] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, "Sequence dependent rigidity of single stranded DNA", Phys. Rev. Lett. **85**, 2400–2403 (2000).

[31] T. Uzawa, T. Isoshima, Y. Ito, K. Ishimori, D. E. Makarov, and K. W. Plaxco, "Sequence and temperature dependence of the end-to-end collision dynamics of single-stranded DNA", Biophys. J. **104**, 2485–2492 (2013).

[32] J. Skolnick and M. Fixman, "Electrostatic persistence length of a wormlike polyelectrolyte", Macromolecules **10**, 944–948 (1977).

[33] J. L. Barrat and J. F. Joanny, "Persistence length of polyelectrolyte chains", Europhys. Lett. **24**, 333+ (2007).

[34] T. Odijk, "Polyelectrolytes near the rod limit", J. Polym. Sci. Polym. Phys. Ed. **15**, 477–483 (1977).

[35] B. Y. Ha and D. Thirumalai, "Electrostatic persistence length of a polyelectrolyte chain", Macromolecules **28**, 577–581 (1995).

[36] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle", Recl. Trav. Chim. Pays-Bas **68**, 1106–1122 (1949).

[37] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules", Proc. Natl. Acad. Sci. U.S.A. **94**, 6185–6190 (1997).

[38] N. M. Toan and D. Thirumalai, "On the origin of the unusual behavior in the stretching of single-stranded DNA", J. Chem. Phys. **136**, 235103 (2012).

[39] M. J. Stevens, D. B. McIntosh, and O. A. Saleh, "Simulations of stretching a strong, flexible polyelectrolyte", Macromolecules **45**, 5757–5765 (2012).

[40] D. Johansen, J. Trewhella, and D. P. Goldenberg, "Fractal dimension of an intrinsically disordered protein: small-angle X-ray scattering and computational study of the bacteriophage λ n protein", Protein Sci. **20**, 1955–1970 (2011).

[41] T. Y. Yoo, S. P. Meisburger, J. Hinshaw, L. Pollack, G. Haran, T. R. Sosnick, and K. Plaxco, "Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state", J. Mol. Biol. **418**, 226–236 (2012).

[42] J. S. Pedersen and P. Schurtenberger, "Scattering functions of semiflexible polymers with and without excluded volume effects", Macromolecules **29**, 7602–7612 (1996).

[43] M. Greenfeld and D. Herschlag, "Probing nucleic acid–ion interactions with buffer exchange-atomic emission spectroscopy", Methods Enzymol. **469**, 375–389 (2009).

[44] Y. Bai, M. Greenfeld, K. J. Travers, V. B. Chu, J. Lipfert, S. Doniach, and D. Herschlag, "Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids", J. Am. Chem. Soc. **129**, 14981–14988 (2007).

[45] W. H. Press, *Numerical recipes 3rd edition: the art of scientific computing* (Cambridge university press, 2007).

[46] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V, and D. I. Svergun, "New developments in the ATSAS program package for small-angle scattering data analysis", J. Appl. Crystallogr. **45**, 342–350 (2012).

[47] T. J. Macke and D. A. Case, "Modeling unusual nucleic acid structures", in *Molecular modeling of nucleic acids*, edited by N. B. Leontis and J. SantaLucia (American Chemical Society, Washington, DC, 1998), pp. 379–393.

[48] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations", Nucleic Acids Res. **32**, W665–W667 (2004).

[49] M. Holst, N. Baker, and F. Wang, "Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples", J. Comput. Chem. **21**, 1319–1342 (2000).

[50] I. Rouzina and V. A. Bloomfield, "Competitive electrostatic binding of charged ligands to polyelectrolytes: practical approach using the non-linear Poisson-Boltzmann equation", Biophys. Chem. **64**, 139–155 (1997).

[51] S. A. Pabit, S. P. Meisburger, L. Li, J. M. Blose, C. D. Jones, and L. Pollack, "Counting ions around DNA with anomalous small-angle X-ray scattering", J. Am. Chem. Soc. **132**, 16334–16336 (2010).

[52] X. Qiu, K. Andresen, L. W. Kwok, J. S. Lamb, H. Y. Park, and L. Pollack, "Inter-DNA attraction mediated by divalent counterions", Phys. Rev. Lett. **99**, 038104+ (2007).

[53] K. Andresen, X. Qiu, S. A. Pabit, J. S. Lamb, H. Y. Park, L. W. Kwok, and L. Pollack, "Mono-and trivalent ions around DNA: a small-angle scattering study of competition and interactions", Biophys. J. **95**, 287–295 (2008).

[54] V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, "Evaluation of ion binding to DNA duplexes using a size-modified Poisson-Boltzmann theory", Biophys. J. **93**, 3202–3209 (2007).

[55] S. Kirmizialtin, A. R. Silalahi, R. Elber, and M. O. Fenley, "The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations", Biophys. J. **102**, 829–838 (2012).

[56] B. I. Kankia, "Binding of Mg^{2+} to single-stranded polynucleotides: hydration and optical studies", Biophys. Chem. **104**, 643–654 (2003).

[57] D. E. Draper, "Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water", Biopolymers **99**, 1105–1113 (2013).

[58] Y. Qu, C. Bolen, and D. Bolen, "Osmolyte-driven contraction of a random coil protein", Proc. Natl. Acad. Sci. U.S.A. **95**, 9268–9273 (1998).

[59] J. Landy, D. McIntosh, and O. Saleh, "Quantifying screening ion excesses in single-molecule force-extension experiments", Phys. Rev. Lett. **109**, 048301 (2012).

[60] A. Dittmore, J. Landy, A. A. Molzon, and O. A. Saleh, "Single-molecule methods for ligand counting: linking ion uptake to dna hairpin folding", J. Am. Chem. Soc. **136**, 5974–5980 (2014).

[61] J. M. Blose, S. A. Pabit, S. P. Meisburger, L. Li, C. D. Jones, and L. Pollack, "Effects of a protecting osmolyte on the ion atmosphere surrounding DNA duplexes", Biochemistry **50**, 8540–8547 (2011).

[62] J. C. Schlatterer, L. W. Kwok, J. S. Lamb, H. Y. Park, K. Andresen, M. Brenowitz, and L. Pollack, "Hinge stiffness is a barrier to RNA folding", J. Mol. Biol. **379**, 859–870 (2008).

[63] S. A. Pabit, J. L. Sutton, H. Chen, and L. Pollack, “Role of ion valence in the submillisecond collapse and folding of a small RNA domain”, Biochemistry **52**, 1539–1546 (2013).

CHAPTER 6

BREAKING THE RADIATION DAMAGE LIMIT WITH CRYO-SAXS

Steve P. Meisburger[1,4], Matthew Warkentin[2,4], Huimin Chen[1], Jesse B. Hopkins[2], Richard E. Gillilan[3], Lois Pollack[1], and Robert E. Thorne[2]

1. Applied and Engineering Physics, Cornell University, Ithaca, New York
2. Department of Physics, Cornell University, Ithaca, New York
3. Cornell High-Energy Synchrotron Source, Ithaca, New York
4. The authors contributed equally to this work

Biophys. J. **104**, 227–36 (2013).

Abstract

Small angle X-ray scattering (SAXS) is a versatile and widely used technique for obtaining low-resolution structures of macromolecules and complexes. SAXS experiments measure molecules in solution, without the need for labeling or crystallization. However, radiation damage currently limits the application of SAXS to molecules that can be produced in microgram quantities; for typical proteins, 10-20 μL of solution at 1 mg/mL is required to accumulate adequate signal before irreversible X-ray damage is observed. Here, we show that cryocooled proteins and nucleic acids can withstand doses at least two orders of magnitude larger than room temperature samples. We demonstrate accurate T = 100 K particle envelope reconstructions from sample volumes as small as 15 nL, a factor of 1000 smaller than in current practice. Cryo-SAXS will thus enable structure determination of difficult-to-express proteins and biologically important, highly radiation sensitive proteins including light-activated switches and metalloenzymes.

6.1 Introduction

In the last decade, small angle X-ray scattering (SAXS) has been developed into a reliable experimental technique for rapidly obtaining low-resolution structures of biological macromolecules. The greatest strength of biological SAXS (BioSAXS) is its ability to report structures from macromolecules in solution, without the need for crystallization or labeling; the macromolecules need only be soluble and in a homogeneous, monodisperse phase. Most synchrotron X-ray sources have dedicated BioSAXS beamlines, including high throughput stations with automated data collection [1–5]. Analysis suites such as ATSAS from EMBL [6] enable rapid and comprehensive interpretation of SAXS data, yielding information ranging from radius of gyration to structural envelopes. This information complements that from atomic resolution techniques and leads to biologically relevant conclusions. SAXS data have been used to validate crystallographic structures and molecular dynamics predictions, to probe conformational switching, to characterize flexibility and folding, and to assemble complexes from atomic structures of subunits [7].

The most important challenges in BioSAXS are to obtain monodisperse, aggregate-free samples, and to maintain this monodispersity throughout X-ray data collection. Radiation damage, which causes aggregation, unfolding and fragmentation, is thus a critical bottleneck. The maximum tolerable X-ray dose (energy per unit mass) is generally orders of magnitude smaller than in X-ray crystallography. For example, lysozyme crystals can withstand $\sim$500 kGy at room temperature [8], whereas lysozyme solutions show excessive aggregation for X-ray doses above $\sim$400 Gy (or $\sim$1 kGy if glycerol is used to modify protein-protein interactions) [9]. Biologically important targets such as metalloproteins and sensors can also exhibit fast damage at specific sites, e.g., at the enzymatically important metal site, that perturb ligand interactions and associated conformation changes. To minimize radiation doses and achieve adequate signal to noise, large sample volumes must be irradiated either by defocusing the X-ray beam [1] or by

flowing [2–4] or translating [10] the sample through the beam. For a typical protein at 1 mg/mL concentration, the minimum sample consumption is roughly 12 μL [1]. Further complicating matters, optimal sample volumes and allowable doses are highly sample-dependent, and must be determined on a case-by-case basis. For example, in SAXS studies of the light-sensing protein VVD, structures of the light-activated and dark states were obtained using a novel coaxial flow cell designed to provide exceptionally short residence times in the X-ray beam, at the cost of increased sample consumption [11].

Radiation damage is also a problem in macromolecular X-ray crystallography (MX) [12–16] and electron microscopy (EM) [17]. In those techniques, radiation damage and minimum sample volumes required for structure determination are dramatically reduced by cooling samples to temperatures near 100 K. Solvent and radical diffusion are all but eliminated, and scaffolding by the frozen solvent network prevents large radiation-induced structural relaxations. In cryo-MX, crystals can withstand a molecule-independent maximum dose of $\sim$30 MGy, 20 to 150 times larger than at room temperature [8, 18–24].

For successful cryocooling, macromolecular structure must be preserved and ice nucleation and growth must be prevented. Solvent vitrification can be achieved by rapid cooling (e.g., by plunging the sample into liquid nitrogen or propane or by inserting in a cold gas stream) [12–16] or by cooling under high pressure [25]. Required cooling rates for complete vitrification can be reduced using chemical cryoprotectants such as glycerol [15, 26, 27]. Although initially developed to reduce radiation damage, sample cryocooling also greatly simplifies storage and dramatically increases shelf-life. It has transformed protein crystallography, enabling high throughput methods such as remote, robotically assisted synchrotron data collection on mailed-in samples [28].

The potential of cryocooling for SAXS studies has long been recognized, but the critical challenge of reproducibly preparing and collecting data from suitable samples has not been

successfully addressed. Unlike MX, SAXS is fundamentally a difference technique. The large contribution of solvent to the total scattering profile from a dilute solution must be measured and subtracted to determine the macromolecule's scattering profile. Scattering from ice crystallites, or from any other electron density inhomogeneity formed during cooling, may be large and irreproducible. Therefore, for cryo-SAXS to become a viable technique, a reliable method for producing homogeneous vitrified samples is required. At the same time, the cryocooling method must preserve the macromolecule's structural integrity and maintain signal to noise in the SAXS profile.

Here, we demonstrate such a method. We have integrated an open-flow nitrogen cryocooler into a standard SAXS beamline, and used SAXS to identify cryoprotection conditions that yield complete vitrification of small drops cooled in the nitrogen gas stream. Using these vitrified samples, we obtain scattering patterns from glucose isomerase (a standard SAXS reference [4, 29, 30]), and verify that low temperatures protect the molecule from radiation damage without altering its structure. Finally, we show that useful cryo-SAXS data can be collected from a variety of macromolecules using very small sample volumes. These methods will be immediately valuable for samples that are especially radiation sensitive, when available sample quantities are limited, or when samples may degrade over time. We discuss remaining challenges that must be overcome for cryo-SAXS to become a widely-adopted alternative to room temperature SAXS for routine measurements of macromolecular structure.

6.2 Materials and Methods

6.2.1 Evaluation of cryoprotectants

The cryoprotectants poly(ethylene glycol) (average molecular mass of 200 Da), glycerol, and dimethyl sulfoxide (DMSO) were purchased from Sigma (St. Louis, MO), and ethylene glycol was purchased from Avantor Performance Materials (Phillipsburg, NJ). Cryoprotectant-water mixtures were prepared by weight in increments of 5 %. Spherical drops of $\sim$1 μL volume were held in a $\sim$700 μm nylon loop (Hampton Research, Aliso Viejo, CA) and rapidly cooled by placing them in a T = 100 K nitrogen cryostream (700 series, Oxford Cryosystems, Oxford, United Kingdom). Cryo-SAXS data on these gas-stream-cooled samples were acquired at CHESS beamline F2. The X-ray energy was 9.88 keV, the sample-detector distance was 1.47 m and the detector was a fiber-coupled CCD (Quantum 1, Area Detector Systems Corporation, Poway, CA). The upstream slits and flight tube were held under vacuum and separated from the sample area by mica windows. SAXS curves were processed using BioXTAS RAW software [31].

6.2.2 Preparation of biological samples for cryo-SAXS

Glucose isomerase crystals (Hampton Research, Aliso Viejo, CA) were re-dissolved in buffer containing 100 mM Tris pH 8.0 and 1 mM $MgCl_2$. Hen egg white lysozyme (Sigma, St. Louis, MO) was dissolved in buffer containing 40 mM Na-acetate pH 4.0, 50 mM NaCl and 1 % (v/v) glycerol. A 24-bp DNA duplex with sequence GGTGACGAGTGAGCTACTGGGCGG (and its complement) was made from synthetic HPLC-purified oligonucleotides (Integrated DNA Technologies, Coralville, IA). The complementary strands were mixed and annealed to form the duplex, following vendor instructions. The DNA was then buffer exchanged with 10 mM Na-

MOPS pH 7.0 and 100 mM NaCl using a spin column (Amicon Ultra-0.5, 10,000 mol wt cutoff, EMD Millipore, Billerica, MA). For each matching buffer, a 2×PEG solution was prepared with 946 mg/mL PEG 200. Each 2×PEG solution was combined with the corresponding cryo-SAXS sample and matching buffer in a 1:1 ratio by volume, for a final concentration of ∼45 % (w/w).

6.2.3 Sample holders

Two different window-free, low volume sample cells were used for cryo-SAXS. For ∼1 μL volume samples, the cell (Fig. 6.2a) was comprised of 1.8 mm long, 860 μm ID, 25 μm wall polyimide tubing. To thermally isolate the sample from the stainless steel support, the sample holder was glued to a short section of 510 μm ID, 25 μm wall polyimide tubing, which was press-fit over the support. The cell was oriented so that the X-ray beam passed along its axis and through the open ends of the tubing. For sub-μL volumes samples, the sample was held by surface tension in a standard polyimide crystallography loop with a 600 μm diameter (MicroMount, MiTeGen), shown in Fig. 6.6a. For room temperature SAXS measurements, an in-vacuum 2 mm quartz capillary with oscillating flow was used to minimize radiation damage [4].

6.2.4 SAXS data collection

SAXS data from biological samples were collected using beamline G1 at CHESS. The beamline was configured with a low-noise area detector (Pilatus 100 K, Dectris, Baden, Switzerland), a He ion chamber for monitoring the incident intensity, and a PIN diode beamstop for measuring transmitted intensity. A cryostream provided nitrogen gas at 100 K for sample cooling (Fig. 6.2a). SAXS data were analyzed using code written in-house in MATLAB (The MathWorks,

Natick, MA).

Because of the small sample volumes used for cryo-SAXS, background scattering originating upstream of the sample had to be minimized. The upstream flight tube was filled with He and extended to within ∼5 mm of the sample using 1.8 mm ID, 0.3 mm wall stainless steel tubing. To further block background scattering that could pass around the sample, the guard slits were supplemented by a 200 nm thick, 500 μm square aperture Si_3N_4 window in a 381 μm thick Si frame (Fabrication Services and Technology Ltd, Northampton, England) that was glued to the end of the tubing. Scattered X-rays were collected through an evacuated flight tube with a 200 nm thick, 2 mm square Si_3N_4 window with (Fabrication Services and Technology Ltd) epoxied to an aluminum cone at the upstream end. The photograph in Fig. 6.2a shows the position of these windows relative to the sample and cryostream.

SAXS data on biomolecules were acquired at CHESS beamline G1 during two separate runs with similar beamline configurations. In the first run, cryo-SAXS data were taken using the 1 μL holder. A 1.52 m sample-detector distance and a 10.5 keV X-ray energy were used to probe scattering wavevectors $0.01 < q < 0.28$ Å^{-1}, where $q = 4\pi\sin(\theta)/\lambda$, 2θ is the scattering angle and λ is the X-ray wavelength. The beam size at the sample position was 119 μm$\times$193 μm (height $\times$ width, FWHM). The X-ray flux calculated from the current through an N_2 ion chamber, placed at the beamstop position with the sample removed, was 6.3×10^{10} s^{-1}. In the second run, cryo-SAXS data from sub-μL samples and room temperature SAXS data were acquired with an X-ray energy of 10.0 keV, a sample-detector distance of 1.35 m, a beam size of 220 μm $\times$190 μm (height $\times$width, FWHM), and an X-ray flux of 1.0×10^{11} s^{-1}. The downstream vacuum window was Kapton film rather than Si_3N_4.

6.2.5 Background subtraction

Conventional SAXS sample cells use parallel, X-ray transparent windows to define a fixed path length. Scattering curves are collected from sample and matching buffer solutions held in identical cells, each is normalized by the transmitted X-ray intensity measured during the exposure, and the two are subtracted to obtain the macromolecule's scattering profile. The cryo-SAXS sample holders described above do not define a fixed path length for the sample, so a normalization and background subtraction method was devised to account for path length variation. The total scattering measured at the detector, $I_{\text{total}}(q)$, is modeled as

$$I_{\text{total}}(q) \propto I_0 T \left\{ \log(1/T) \left[I_{\text{M}}(q) + I_{\text{S}}(q) \right] + I_{\text{bkg}}(q) \right\} \tag{6.1}$$

where I_0 is the incident intensity, T is the X-ray transmission factor of the sample (and thus $\log(1/T)$ is proportional to the thickness), $I_{\text{M}}(q)$ and $I_{\text{S}}(q)$ are the scattering from the macromolecule and solvent, respectively, and $I_{\text{bkg}}(q)$ is the instrumental background scattering. With appropriate normalization, $I_{\text{M}}(q)$ can be found from three scattering profile measurements: one of the macromolecule-containing sample, one of the macromolecule-free buffer, and one of the empty cell. The incident and transmitted intensities, I_{incident} and $I_{\text{transmitted}}$, are measured at the same time as $I_{\text{total}}(q)$, and T is calculated using

$$T = \left(I^{\text{empty}}_{\text{incident}} / I^{\text{empty}}_{\text{transmitted}} \right) \times \left(I_{\text{transmitted}} / I_{\text{incident}} \right) \tag{6.2}$$

where the first factor, obtained from measurements of the empty cell, is included to cross-calibrate the two detectors. A background subtracted, thickness normalized intensity can be computed for both the sample (macromolecule plus buffer) and the buffer. Here background subtraction refers to removing the properly scaled instrumental background scattering:

$$I_{\Delta}(q) \equiv \left(\frac{I_{\text{total}}(q)}{I_{\text{transmitted}}} - \frac{I^{\text{empty}}_{\text{total}}(q)}{I^{\text{empty}}_{\text{transmitted}}} \right) \frac{1}{\log(1/T)} \tag{6.3}$$

Then, the difference between buffer and sample curves $I_{\text{diff.}}(q) = I^{\text{sample}}_{\Delta}(q) - I^{\text{buffer}}_{\Delta}(q)$ is proportional to $I_{\text{M}}(q)$, the scattering of the macromolecule.

6.2.6 Estimation of X-ray dose

The X-ray dose D delivered to the sample was calculated using

$$D = \frac{t_{exp} f E (1-T)}{V \rho} \tag{6.4}$$

where t_{exp} is the exposure time, f is the X-ray flux (photons per second), E is the X-ray energy, T is the sample transmission factor, V is the illuminated volume, and ρ is the mass density. The density $\rho \approx 1.07\,\mathrm{g\,cm^{-3}}$ was estimated from available data on PEG-water mixtures at room temperature [32]. For each sample, the illuminated volume was found from $V \approx A\,\mu^{-1}\log(1/T)$ where A is the beam area (product of width and height at FWHM) and μ^{-1} is the X-ray absorption length. The absorption length was calculated from atomic absorption data [33] and by approximating the 45 % (w/w) PEG-200 water mixture as tetra(ethylene glycol) and water in a 1:13 molar ratio (i.e. $H_{44}O_{18}C_8$) with the density given above; at 10 keV, $\mu^{-1} \approx 2.15$ mm. In X-ray crystallography, dose calculations are typically based on the mass-energy absorption coefficient [34]. At the X-ray energies used, this calculation agrees with one based on the mass-energy absorption coefficient to within 1 %.

An accurate measure of the beam size at the sample position is required for computing the X-ray dose. X-ray burns in a 1 mm thick glass slide were acquired with exposures of 1, 2, 8, 16, and 32 seconds, digitized using a flatbed scanner at 12,800 dpi (EPSON Perfection 1660), and analyzed in MATLAB. The images were corrected for the nonlinear response of the glass using a calibration curve generated from the multiple exposures.

6.2.7 Analysis of SAXS profiles

The pair-distance distribution function P(r) was calculated from the SAXS profiles using the Bayesian Indirect Fourier Transform (BIFT) method [35]. An indirect Fourier transform pro-

gram with smoothness regularization was written in MATLAB, and Bayesian estimation was used to find the maximum particle dimension D_{max}, the Lagrange multiplier α, and the noise level β [35]. In addition, the evidence for the hypothesis was computed (i.e., the probability of the data given the basis set, noise model, and regularizer) [36].

Ab initio reconstructions of the low-resolution particle envelope from SAXS data were performed using the ATSAS suite of programs [6]. For each SAXS curve, 16 DAMMIF reconstructions (fast mode, no symmetry) were aligned and averaged using DAMAVER. Each average reconstruction was aligned with its corresponding atomic structure using SUPCOMB and visualized using Pymol version 1.2r1 (DeLano Scientific LLC).

6.3 Results and Discussion

6.3.1 PEG-200 solutions yield good contrast, low background scattering and complete vitrification of 1 μL drops

Excessive and irreproducible ice formation on cooling has been a major obstacle to cryo-SAXS. In cryo-MX [37–39], the cooling rate and the choice and concentration of cryoprotectant are key variables in obtaining a homogeneous, fully vitrified state at T = 100 K. Cryo-SAXS imposes additional constraints. Most cryoprotectants have higher electron densities than water. Adding cryoprotectant thereby increases the solvent's average electron density and its electron density fluctuations, decreasing SAXS contrast and increasing solvent background scattering. Consequently, cryoprotectants that are effective at low concentrations and that have electron densities near that of water are preferred. Based upon these criteria, we find PEG-200 to be the superior choice among several other small-molecule cryoprotectants tested (including glycerol, ethylene

glycol, and DMSO).

The cryoprotectant concentration necessary for complete vitrification was determined by acquiring cryo-SAXS profiles from $\sim$1 μL PEG-water drops at several PEG-200 concentrations (Fig. 6.1, b and c). At concentrations below 45 % (w/w), the presence of ice is indicated by a steep rise in the scattered intensity below $q \approx 0.02$ Å^{-1}, in some cases to almost four orders of magnitude above the profile's high-q baseline. When plotted on a log-log scale, it can be seen that the signature of ice formation is a power law at all concentrations (Fig. 6.1d). At 45 % (w/w) PEG, the low-q scattering is 10 times the baseline level in one sample (*G*), and absent in a second sample (*H*). In photographs of the sample drops (Fig. 6.1a), samples *F* and *G*, with PEG concentrations of 40 % and 45 % (w/w), respectively, are both visually clear. Thus, sample clarity is not a sufficient indicator of complete vitrification in the context of cryo-SAXS.

At PEG concentrations above 45 % (w/w), no ice signal was observed, and samples could be reliably vitrified. Similar measurements yielded minimum concentrations for ice-free cryo-SAXS profiles of 50 % (w/w) for glycerol, 50 % (w/w) for ethylene glycol, and 45 % (w/w) for DMSO. These cryoprotectant concentrations are roughly 5 % larger than are required to eliminate ice rings in crystallographic diffraction at comparable cooling rates [26].

6.3.2 Sample cell design and buffer subtraction technique enable collection of cryo- SAXS data from biomolecules

The window-free, thin-wall sample cell was designed to optimize the X-ray path length through the 1 μL sample for good signal to noise, and to maximize heat transfer rates through its side-walls for rapid cooling (Fig. 6.2a). The X-ray beam passed along the axis of the cylindrical cell, and the sample was held within it by surface tension prior to cooling. At the Cornell

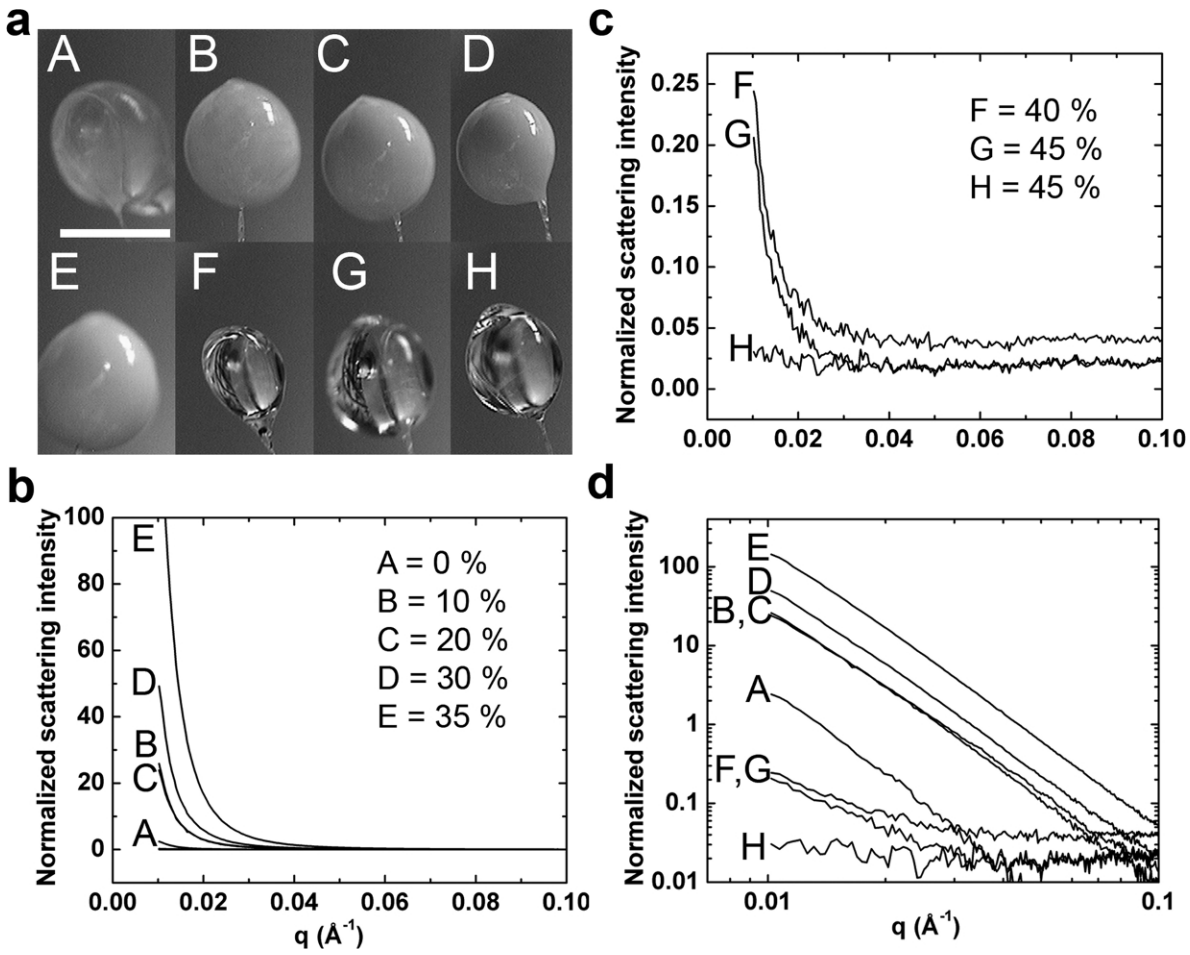

Figure 6.1: Cryoprotectant concentration required to obtain an ice-free scattering curve. (a) Photographs of $\sim$1 μL drops of PEG 200 - water mixtures held in $\sim$700 μm nylon loops after cooling to 100 K in a N_2 gas stream. PEG concentrations for drops (A-H) ranged from 0 % to 45 % (w/w) (indicated in the legends of (b) and (c)). Opacity of the drop arises from light scattering by ice crystals. (b) SAXS curves show increasing scattering intensity at $q \lesssim 0.02$ Å^{-1} as the PEG concentration increases from 0 to 35 % (w/w). (c) At 40 % (w/w) PEG and above the drops are visually clear and the $q \lesssim 0.02$ Å^{-1} intensity drops dramatically. However, scattering at low q values still shows the presence of small amounts of ice. Using 45 % (w/w) (drop H), the excess low-q scattering is absent. (d) The scattering curves from (b) and (c) are shown on log-log axes, revealing the characteristic power-law scattering of ice crystallites at all concentrations.

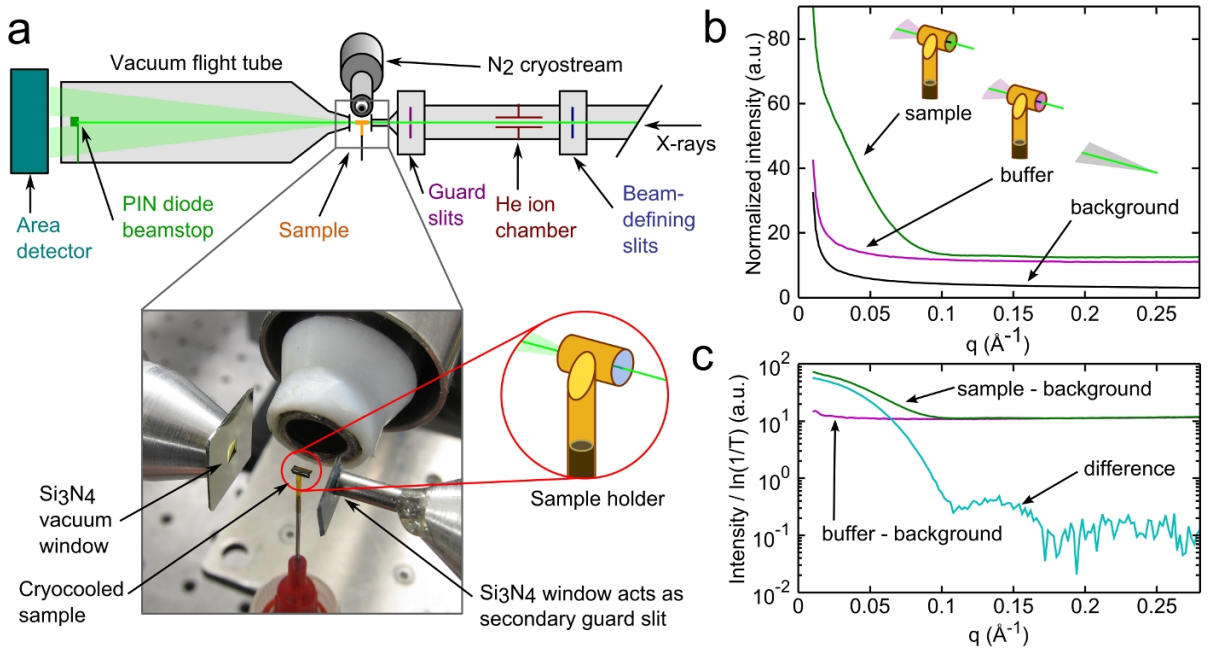

Figure 6.2: Apparatus and method for obtaining SAXS profiles from solution samples at 100 K. (a) A SAXS beamline was configured with a cryostream providing a steady flow of T = 100 K N_2 gas at the sample position. Monitors for incident intensity (ion chamber) and transmitted intensity (PIN diode) allowed measurement of the X-ray transmission factor for each sample and normalization of scattering profiles. X-ray windows of 200 nm thick Si_3N_4 on the upstream and downstream flight tubes minimized background scattering. The 500 μm square aperture of the upstream window functioned as a second guard slit. The window-free sample holder was made from a 1.8 mm section of thin-wall polyimide tubing. A $\sim$1 μL sample was injected into the tubing and then vitrified by cooling in the gas stream. The sample cell was oriented so that the X-ray beam passed along the central axis of the tubing. The photograph shows a sample in the cryostream. (b) Cryo-SAXS profiles obtained from vitrified solutions of 2 mg/ml glucose isomerase (GI) and its matching buffer, as well as for the instrumental background with the sample removed. Slight differences in the X-ray path lengths through the GI and buffer samples contributed to differences in their scattering. (c) Measurement of the X-ray transmission factor, T, and normalization by the path-length, $\log(1/T)$, provided accurate background subtraction using the data in (b) to obtain GI's SAXS profile.

High Energy Synchrotron Source (CHESS), a cryostream cooler was incorporated into the G1 beamline's SAXS setup, allowing sample cooling by a continuous nitrogen gas stream at a temperature of 100 K.

In conventional SAXS on liquid samples, the fixed sample cell windows precisely define the X-ray path length, allowing buffer subtraction of data acquired in the same cell. In our window-

free cell, the path length depends on the sample's volume and the shape of its meniscus. A three-curve background subtraction method was used to correct for inevitable path length variations, as described in Materials and Methods. To demonstrate this method, Fig. 6.2b shows cryo-SAXS profiles for a 2 mg/mL glucose isomerase (GI) solution and its matching buffer, as well as the instrumental background. Each solution contained the cryoprotectant PEG-200 at 45 % (w/w) concentration. In Fig. 6.2c, subtraction of the normalized and background subtracted GI and buffer solution profiles reveals the small oscillations at high-q that are characteristic of a large, sphere-like globular protein.

6.3.3 Radius of gyration, maximum dimension, and particle envelope determined by cryo-SAXS for glucose isomerase

Macromolecular SAXS data are most often used to find shape information, including the radius of gyration (R_g), the maximum particle dimension, and the low resolution envelope. To assess whether cryo-SAXS data are of sufficient quality for these purposes, the cryo-SAXS profile for 2 mg/ml GI was analyzed using standard techniques [40]. The Guinier plot shown in Fig. 6.3a is linear within the noise down to the smallest angles measured ($q = 0.01$ Å^{-1}), and the slope gives a radius of gyration $R_g = 33.4 \pm 0.1$ Å in excellent agreement with the crystal structure-derived value of 33.35 Å (CRYSOL with default parameters [6] and PDB 1XIB). The pair-distance distribution function P(r) in Fig. 6.3b has a Gaussian shape characteristic of a globular particle, and gives a well-defined maximum dimension. An ab initio reconstruction of the molecular envelope without symmetry constraints shows fair agreement with the tetrameric structure from MX (Fig. 6.3c); imposing appropriate symmetry constraints during reconstruction improves the agreement (Fig. C.1).

To determine whether the cryoprotectant or the cooling process significantly alters the SAXS

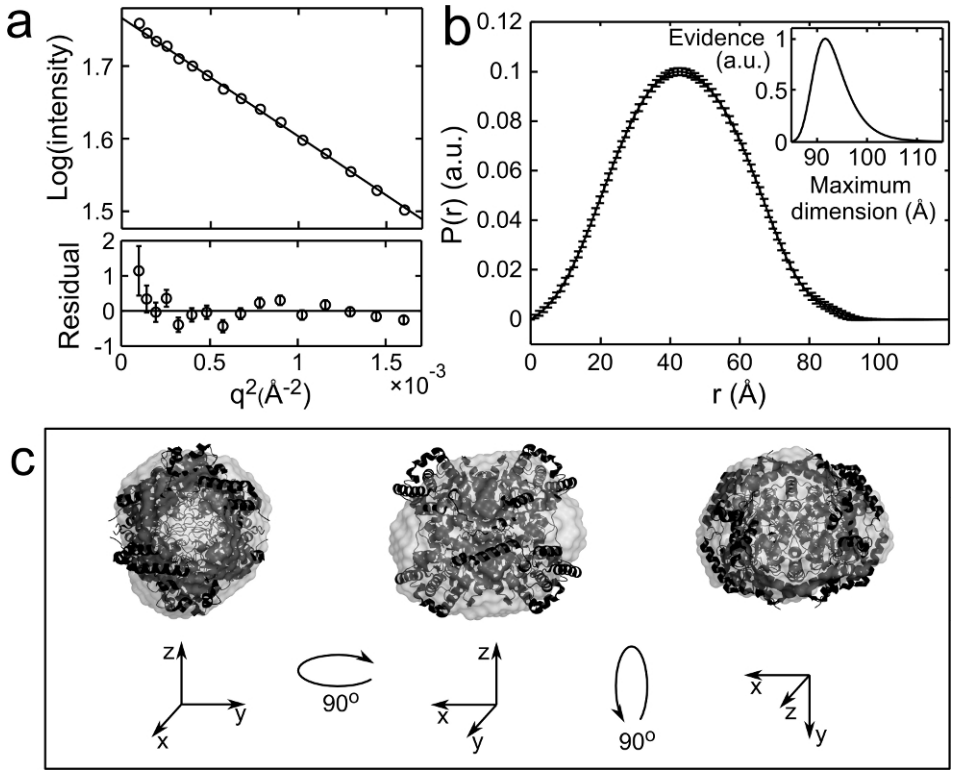

Figure 6.3: Analysis of cryo-SAXS data from glucose isomerase. (a) Guinier plot with linear fit used to find the radius of gyration R_g. (b) The pair distance distribution function P(r) derived from the Fourier transform of the scattering profile provides information about the particle shape including the maximum particle dimension (inset). (c) Three orientations of the reconstructed particle envelope with the docked crystal structure. The mean NSD was 0.641.

profile, data were collected at room temperature (RT) from GI in buffer with and without 45 % (w/w) PEG-200 cryoprotectant. The samples were oscillated through the beam within a stationary quartz capillary to reduce radiation damage. The most obvious effect of PEG-200 is a reduction in signal intensity due to lower electron density contrast between protein and solvent. Assuming average electron densities of 0.334 $Å^{-3}$ for water, 0.420 $Å^{-3}$ for protein [41], and 0.355 $Å^{-3}$ for a 45 % (w/w) PEG-200 water solution at 298 K [32], the scattering of protein in 45 % (w/w) PEG should be 57 % of its scattering in pure water. This is consistent with the observed two-fold reduction in scattering intensity of GI in PEG compared with standard buffer at room temperature. Despite this change in intensity, the overall shape of the scattering curve for GI is otherwise unchanged (Fig. 6.4). R_g determined from the Guinier plot was 32.8 ± 0.1 Å

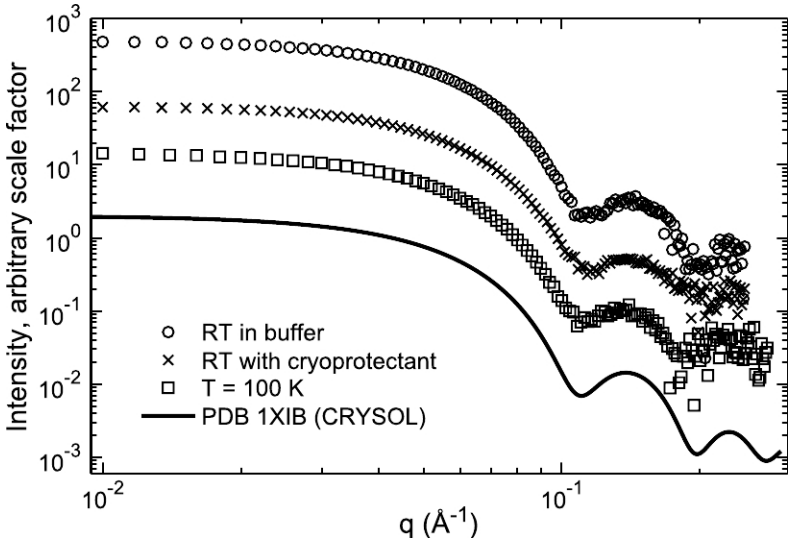

Figure 6.4: Background subtracted and normalized SAXS profiles from 2 mg/ml glucose isomerase at room temperature and 100 K. The presence of cryoprotectant (45% w/w PEG-200) and rapid cooling to 100 K do not affect the essential features of the scattering profile. The CRYSOL prediction from the crystal structure of GI was calculated using default parameters, and is shown for reference. For display purposes, curves were multiplied by an arbitrary scale factor. The total exposure times for room temperature samples were 32 s in aqueous buffer and 52 s in PEG buffer. The sample at T = 100 K was exposed for 180 s (corresponding to a dose of 220 kGy).

in buffer and 32.5 ± 0.1 Å with cryoprotectant added. These values agree within experimental error with the average R_g of 32.7 ± 0.2 Å determined in previous SAXS studies of GI [29].

The effect of cryo-cooling on GI's scattering profile is minimal. The profile's shape and calculated R_g are similar to those from room temperature measurements and from the crystal structure prediction. Small differences between the 100 K and room temperature SAXS curves are observed in the Guinier region (Fig. C.2). The calculated R_g is $\sim$2 % larger at 100 K. This difference is small compared with experimental uncertainties, and may arise from differences in the sample geometry for 100 K and room temperature measurements, and also from residual interparticle interference at the working concentration of 2 mg/mL (Fig. C.3). However, since a protein's SAXS profile includes scattering from a hydration layer of ordered water molecules, the 2 % increase in R_g at 100K may reflect increased hydration water ordering, as is seen in

protein crystals [42].

6.3.4 SAXS from cryocooled glucose isomerase is insensitive to large X-ray doses

X-ray induced changes to the macromolecule's structure or solution state must be minimized to obtain reliable SAXS profiles. In room temperature solution SAXS, a series of profiles are acquired and inspected for dose and time-dependent changes using the radius of gyration as a means of quantifying damage and determining the maximum tolerable X-ray exposure [9, 10]. Collecting data on vitrified samples at T $\sim$100 K should eliminate radiation induced aggregation (which otherwise dominates low-angle scattering) and reduce unfolding and fragmentation.

To verify that data collection at T = 100 K reduces the rate at which radiation damage is manifested in SAXS profiles, a series of 60 s exposures were acquired from a $\sim$1 μL GI + buffer sample and then from a buffer-only sample. The accumulated dose for each exposure was calculated from the incident X-ray beam intensity and illuminated volume as described in Materials and Methods. GI's SAXS profile at a given dose was obtained by subtracting a buffer curve at roughly the same dose. At all doses, the R_g values fall between 33 and 34 Å and do not show any obvious dose-dependence (Fig. 6.5a). The SAXS profiles obtained from the first and last exposures are indistinguishable (Fig. 6.5b). The final exposure corresponds to an accumulated dose of 3.7 MGy.

In cryoelectron microscopy and diffraction, in X-ray cryocrystallography, and in X-ray diffractive imaging of biological samples, all proteins show similar radiation sensitivity – measured on a damage per dose basis – at T = 100 K [21, 43, 44], and this should also be true in cryo-SAXS. For these other diffraction techniques, the maximum tolerable dose at T = 100 K

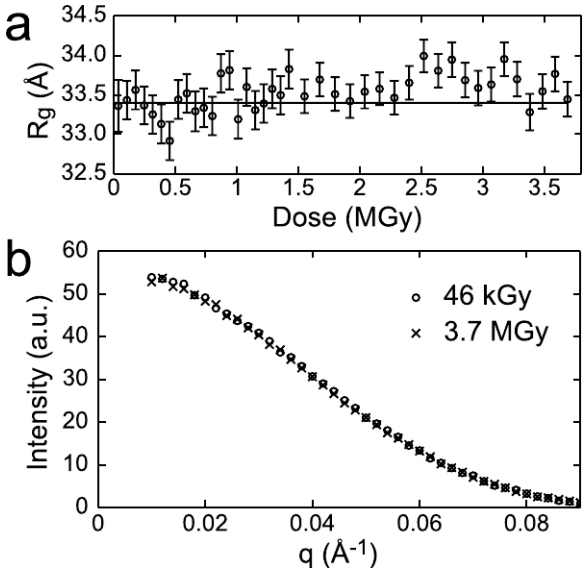

Figure 6.5: The SAXS profile of glucose isomerase is insensitive to large X-ray doses. (a) Radius of gyration (R_g) of GI as a function of accumulated X-ray dose for a single sample shows no radiation damage even at high doses. R_g was calculated from the region $0.02 < q < 0.04$ $\mathring{A}^{-1}$, and standard errors computed from the fits are shown. (b) The first and last SAXS profiles in the dose series of (a) are indistinguishable.

to achieve a data set of a given resolution (in Å) is roughly 10 MGy/Å [44, 45]. For a SAXS data set to $q = 0.3$ $\mathring{A}^{-1}$, corresponding to a resolution of $\sim$20 Å, this yields a maximum tolerable dose of 200 MGy. However, analysis of SAXS data is sensitive both to loss of information at high-q and to radiation-induced changes at low-q. At room temperature, radiation damage first manifests at low-q, presumably because of molecular aggregation, fragmentation and unfolding, processes that should be strongly suppressed at T = 100 K. However, at large doses, microscopic inhomogeneities due to, for example, radiolytic cleavage of hydrogen and subsequent recombination and diffusion, may develop. In cryoelectron microscopy, hydrogen bubbles become evident beyond doses of $\sim 1000 - 10,000$ electrons/nm^2 [46] corresponding to doses of $\sim$45 MGy [47]. Recent SAXS measurements on cryo-cooled insulin crystals observed a strong increase in scatter attributed to hydrogen bubble formation beyond $\sim$70 MGy, increasing to >180 MGy for T$\leq$30 K [48]. Consequently, for typically radiation sensitive biomolecules with room temperature dose limits of 1-10 kGy, cryo-SAXS should yield dose limit increases of

between two and five orders of magnitude, and corresponding reductions in minimum sample volumes.

6.3.5 High dose tolerance enables macromolecular envelope reconstruction from nL sample volumes

Because of the greatly increased dose tolerance at T = 100 K, cryo-SAXS should enable dramatic reductions in minimum sample volumes for macromolecular envelope reconstruction. To demonstrate this, cryo-SAXS data were collected from non-spherical drops with thicknesses between 300 and 500 μm, held within polyimide crystallography loops with a 600 μm diameter (Fig. 6.6a). The X-ray illuminated volume (the product of the drop thickness and beam area) ranged from 13 to 25 nL.

With this non-ideal sample geometry, accurate buffer subtraction at high-q was more difficult to achieve than with the sample holders of Fig. 6.2. In general, SAXS patterns from macromolecules decay rapidly toward background at high q. This makes the high-q background subtraction exquisitely sensitive to errors in the normalization of sample and buffer curves. Although we were unable to identify sources of these errors, SAXS profiles of multiple cryocooled drops containing the same PEG-buffer solution have subtraction errors that are well approximated within noise by a constant offset (Fig. C.4).

To correct for these background subtraction errors in sub-μL drops, constant offsets were chosen to maximize the so-called Bayesian evidence of the P(r) function [50] obtained using the Bayesian Indirect Fourier Transform (BIFT) method. The implementation of BIFT is described in Materials and Methods, and its ability to correct offset errors under experimental conditions is demonstrated with synthetic data in Fig. C.5. In general it is not good practice to add offsets

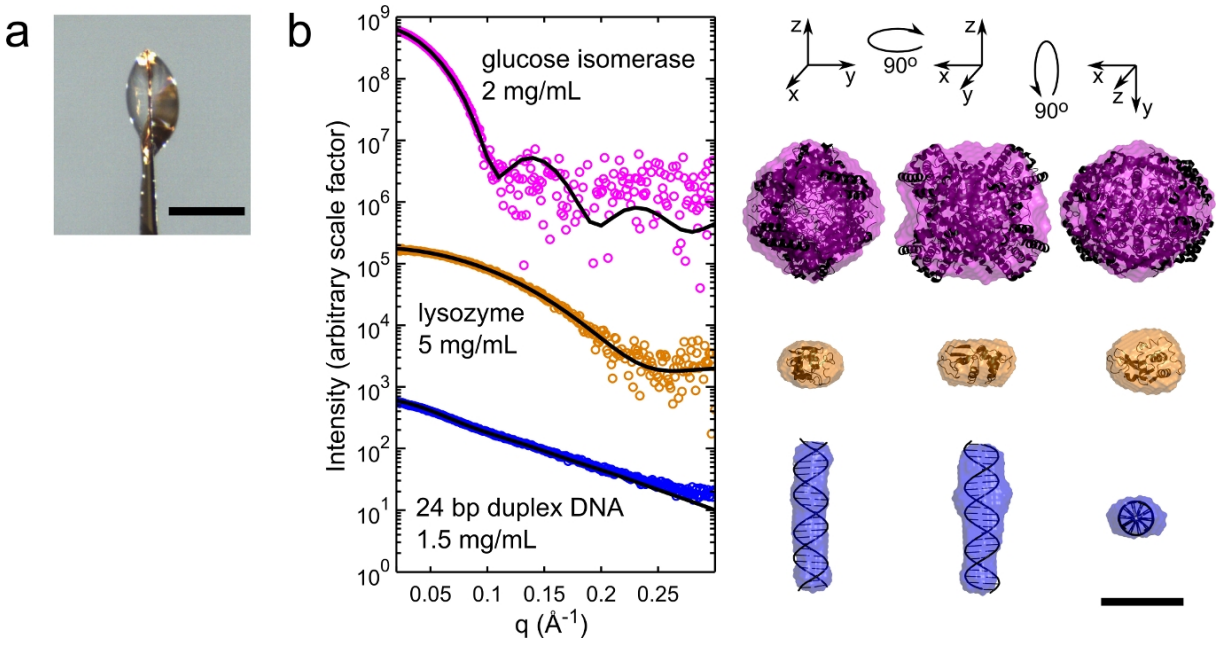

Figure 6.6: Molecular envelopes from nanoliter volumes. (a) Small, lenticular sample drops were held in a 600 μm diameter polyimide crystallography mount. A representative drop is shown. The scale bar in the image is 500 μm. (b) Cryo-SAXS data were acquired for glucose isomerase (GI), hen egg white lysozyme, and 24 base-pair duplex DNA at the indicated concentrations. Cryo-SAXS profiles for each macromolecule were corrected by applying a constant offset using the BIFT method, as described Materials and Methods (this constant offset arises from sample geometry-dependent background subtraction errors at high-q values). Solid lines show the CRYSOL predictions from each atomic structure, based on PDBs 1XIB and 2LYZ, and an ideal 24-bp DNA helix generated using Nucleic Acid Builder [49]. No fitting parameters were used except for an overall scale factor for the (arbitrary) intensity. Macromolecule envelopes generated from the cryo-SAXS data were aligned with the atomic structures, and are shown in three orientations. All are scaled according the 50 Å bar at lower right. For GI, lysozyme and DNA, the mean NSDs were 0.586, 0.459 and 0.542; the X-ray illuminated sample volumes were 16.3, 13.6 and 24.9 nL; the exposure times were 160, 80, and 160 seconds; and the X-ray doses were 275, 114, and 234 kGy, respectively.

to SAXS profiles, as the high-q scattering is used to determine the foldedness (Kratky plot) and volume (Porod invariant) of the macromolecule. However, this correction can be done without biasing envelope reconstructions. For example, the program DAMMIN/F of the ATSAS suite by default adjusts the constant offset to account for the bead model's inability to accurately represent internal density fluctuations [51].

Fig. 6.6b shows cryo-SAXS profiles and calculated molecular envelopes for GI, lysozyme,

and 24-bp DNA duplex. Each profile was determined from one sample drop and one buffer drop, with an X-ray dose to the nanoliter samples of between 100 and 300 kGy. The profiles are noisy, but in all cases are sufficient to obtain molecular envelopes in reasonable agreement with atomic structures determined from crystallography.

6.4 Conclusions

The results presented here demonstrate the basic feasibility of cryo-SAXS as a method for determining structural information from macromolecules in sub-μL volume samples. Macromolecular solutions can be cooled into a vitrified state exhibiting no excess low-q scatter, indicating the absence of ice or other inhomogeneities on the length scales probed by SAXS. Required cryoprotectant concentrations are tolerable, produce modest reductions in SAXS contrast, and (at least for the macromolecules studied here) do not affect macromolecule structure. Because aggregation, unfolding, fragmentation and other degradation processes that generate sample inhomogeneities are largely eliminated in vitrified samples, radiation damage per unit dose is reduced by at least two and as much as five orders of magnitude relative to room-temperature SAXS; unless radiation-induced hydrogen bubble formation becomes important, cryo-SAXS dose limits should be $\sim$100 MGy, substantially larger than in cryocrystallography because of the lower resolution provided by SAXS [44]. This large decrease in radiation sensitivity allows minimum sample volumes to be reduced by a comparable factor. Solvent and instrumental background subtraction is possible even with non-ideal and non-identical macromolecule and buffer sample geometries, allowing determination of radii of gyration and structural envelopes that match results from crystallography.

While elimination of ice crystallites allowed us to subtract cryo-SAXS buffer scattering at low-q, reliable buffer subtraction at high-q is still a challenge. Sample geometry nonidealities

and irreproducibilities (absent from the fixed, parallel-wall cell geometry used in room temperature SAXS) and drift in instrumental background introduce errors. However, by configuring the SAXS beamline to minimize background, designing a sample cell to optimize signal to noise, and measuring and correcting for path length variations, we were able to demonstrate reasonable buffer subtraction for 2 mg/mL GI. Future application of cryo-SAXS to weakly scattering systems such as low MW proteins, dilute solutions, and small sample volumes, may require additional optimization of the technique. For small sample volumes, we found that background subtraction errors were present, but could be corrected analytically. This analytical correction is adequate for generating molecular envelopes, but it is not ideal for general SAXS work. More experiments must be done to characterize the important sources of error and to address them through engineering of sample holders and beamline hardware.

The primary effect of the 45% (w/w) PEG-200 concentration used here was a $\sim$two-fold reduction in macromolecule contrast. However, cryoprotectants may also have effects on macromolecular structure and interactions. Cryoprotectants are osmolytes, and therefore modify water activity and macromolecule hydration. High osmolyte concentrations have been used to mimic cellular conditions in vitro, and therefore may help preserve biologically relevant solution structures over some concentration range. Glycerol is known to stabilize protein structure and prevent aggregation [52]. PEG-200 decreases the melting temperature of nucleic acid secondary structures, but stabilizes tertiary structures [53–56]. These and other effects of cryoprotectant will have to be taken into account when comparing cryo-SAXS results with those from other techniques. Smaller sample volumes and the use of liquid nitrogen or propane instead of gaseous nitrogen as the cooling agent should allow cryoprotectant concentrations to be decreased below 30 % [38, 57, 58], comparable to values routinely used in cryo-MX and generally assumed to have negligible effects on macromolecule structure.

The use of cryocooling has the potential to eliminate many difficulties associated with room

temperature SAXS experiments. Because vitrified samples are much more radiation tolerant, they may yield much larger integrated signal to noise with no concerns about damage. Samples that spontaneously aggregate or otherwise degrade with time can be frozen immediately after manufacture and stored indefinitely, eliminating concerns about long-term stability. Dramatically reduced sample volume requirements will facilitate mass screening of solution conditions for their effects on molecular structure or association, and combinatorial binding assays to, e.g., elucidate pathways for macromolecular complex formation. Short turn around times will result from exploiting the existing infrastructure for high-throughput crystallography, including mail-in facilities that employ automated sample handling and full brightness, for rapid turn-around data collection. In the same way that cryo-MX has transformed atomic resolution studies, cryo-SAXS is poised to transform low-resolution studies of macromolecular structure and function.

Acknowledgments

We thank Arthur Woll and CHESS staff for experimental assistance and Thorne and Pollack group members for help with synchrotron data collection.

R.E.T. acknowledges support from the NIH (R01 GM65981) and the NSF (DMR-0805240). L.P. acknowledges support from NIH (R01-GM085062). CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-0936384, and the MacCHESS resource is supported by NIGMS award GM-103485. R.E.T. acknowledges a significant financial interest in MiTeGen, LLC which manufactured some of the tools used in this work.

References

[1] G. L. Hura, A. L. Menon, M. Hammel, R. P. Rambo, F. L. P. Ii, S. E. Tsutakawa, F. E. J. Jr, S. Classen, K. A. Frankel, R. C. Hopkins, S.-j. Yang, J. W. Scott, B. D. Dillard, M. W. W. Adams, and J. A. Tainer, "Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)", Nat. Methods **6**, 606–612 (2009).

[2] P. Pernot, P. Theveneau, T. Giraud, R. Nogueira Fernandes, D. Nurizzo, D. Spruce, J. Surr, and S. McSweeney, "New beamline dedicated to solution scattering from biological macromolecules at the ESRF", J. Phys: Conf. Ser. **247**, 012009 (2010).

[3] A. Martel, P. Liu, T. M. Weiss, M. Niebuhr, and H. Tsuruta, "An integrated high-throughput data acquisition system for biological solution X-ray scattering studies", J. Synchrotron Radiat. **19**, 431–434 (2012).

[4] S. S. Nielsen, M. Møller, and R. E. Gillilan, "High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell", J. Appl. Crystallogr. **45**, 213–223 (2012).

[5] C. Blanchet, A. Zozulya, A. Kikhney, D. Franke, P. Konarev, W. Shang, R. Klaering, B. Robrahn, C. Hermes, F. Cipriani, et al., "Instrumental setup for high-throughput small- and wide-angle solution scattering at the X33 beamline of EMBL Hamburg", J. Appl. Crystallogr. **45**, 489–495 (2012).

[6] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V, and D. I. Svergun, "New developments in the ATSAS program package for small-angle scattering data analysis", J. Appl. Crystallogr. **45**, 342–350 (2012).

[7] R. P. Rambo and J. A. Tainer, "Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering", Curr. Opin. Struct. Biol. **20**, 128–137 (2010).

[8] R. J. Southworth-Davies, M. A. Medina, I. Carmichael, and E. F. Garman, "Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography", Structure **15**, 1531–1541 (2007).

[9] S. Kuwamoto, S. Akiyama, and T. Fujisawa, "Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants", J. Synchrotron Radiat. **22**, 462–468 (2004).

[10] X. Hong and Q. Hao, "Measurements of accurate X-ray scattering data of protein solutions using small stationary sample cells", Rev. Sci. Instrum. **80**, 014303 (2009).

[11] J. S. Lamb, B. D. Zoltowski, S. A. Pabit, L. Li, B. R. Crane, and L. Pollack, "Illuminating solution responses of a LOV domain protein with photocoupled small-angle X-ray scattering", J. Mol. Biol. **393**, 909–19 (2009).

[12] H. Hope, "Cryocrystallography of biological macromolecules: a generally applicable method", Acta Crystallogr. B **44**, 22–26 (1988).

[13] H. Hope, "Crystallography of biological macromolecules at ultra-low temperature", Annu. Rev. Biophys. Biophys. Chem. **19**, 107–126 (1990).

[14] D. W. Rodgers, "Cryocrystallography", Structure **2**, 1135–1140 (1994).

[15] E. Garman and T. Schneider, "Macromolecular cryocrystallography", J. Appl. Crystallogr. **30**, 211–237 (1997).

[16] J. W. Pflugrath, "Macromolecular cryocrystallography–methods for cooling and mounting protein crystals at cryogenic temperatures", Methods **34**, 415–423 (2004).

[17] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz, "Cryo-electron microscopy of vitrified specimens", Q. Rev. Biophys. **21**, 129–228 (1988).

[18] W. Hendrickson, "Radiation damage in protein crystallography", J. Mol. Biol. **106**, 889–893 (1976).

[19] T. Y. Teng and K. Moffat, "Primary radiation damage of protein crystals by an intense synchrotron X-ray beam", J. Synchrotron Radiat. **7**, 313–317 (2000).

[20] T. Y. Teng and K. Moffat, "Radiation damage of protein crystals at cryogenic temperatures between 40 K and 150 K", J. Synchrotron Radiat. **9**, 198–201 (2002).

[21] R. L. Owen, E. Rudiño-Piñera, and E. F. Garman, "Experimental determination of the radiation dose limit for cryocooled protein crystals", Proc. Natl. Acad. Sci. U.S.A. **103**, 4912–4917 (2006).

[22] J. Kmetko, N. S. Husseini, M. Naides, Y. Kalinin, and R. E. Thorne, "Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures", Acta Crystallogr. D Biol. Crystallogr. **62**, 1030–1038 (2006).

[23] J. Kmetko, M. Warkentin, U. Englich, and R. E. Thorne, "Can radiation damage to protein crystals be reduced using small-molecule compounds?", Acta Crystallogr. D Biol. Crystallogr. **67**, 881–893 (2011).

[24] M. Warkentin and R. E. Thorne, "Glass transition in thaumatin crystals revealed through temperature-dependent radiation-sensitivity measurements", Acta Crystallogr. D Biol. Crystallogr. **66**, 1092–1100 (2010).

[25] C. U. Kim, R. Kapfer, and S. M. Gruner, "High-pressure cooling of protein crystals without cryoprotectants", Acta Crystallogr. D Biol. Crystallogr. **61**, 881–890 (2005).

[26] M. B. McFerrin and E. H. Snell, "The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images", J. Appl. Crystallogr. **35**, 538–545 (2002).

[27] D. H. Juers and B. W. Matthews, "Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization", Q. Rev. Biophys. **37**, 105–119 (2004).

[28] H. Robinson, A. S. Soares, M. Becker, R. Sweet, and A. Héroux, "Mail-in crystallography program at Brookhaven National Laboratory's National Synchrotron Light Source", Acta Crystallogr. D Biol. Crystallogr. **62**, 1336–1339 (2006).

[29] M. Kozak, "Glucose isomerase from Streptomyces rubiginosus – potential molecular weight standard for small-angle X-ray scattering", J. Appl. Crystallogr. **38**, 555–558 (2005).

[30] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution", Q. Rev. Biophys. **40**, 191–285 (2007).

[31] S. S. Nielsen, K. N. Toft, D. Snakenborg, M. G. Jeppesen, J. K. Jacobsen, B. Vestergaard, J. P. Kutter, and L. Arleth, "BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis", J. Appl. Crystallogr. **42**, 959–964 (2009).

[32] E. A. Muller and P. Rasmussen, "Densities and excess volumes in aqueous poly(ethylene glycol) solutions", J. Chem. Eng. Data **36**, 214–217 (1991).

[33] B. L. Henke, E. M. Gullikson, and J. C. Davis, "X-ray interactions: photoabsorption, scattering, transmission, and reflection at E = 50-30,000 eV, Z = 1-92", Atom. Nucl. Data Tables **54**, 181–342 (1993).

[34] K. S. Paithankar and E. F. Garman, "Know your dose: *raddose*", Acta Crystallogr. D Biol. Crystallogr. **66**, 381–388 (2010).

[35] S. Hansen, "Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering", J. Appl. Crystallogr. **33**, 1415–1421 (2000).

[36] D. MacKay, "Bayesian interpolation", Neural Comput. **4**, 415–447 (1992).

[37] U. Chinte, B. Shah, K. DeWitt, K. Kirschbaum, A. A. Pinkerton, and C. Schall, "Sample size: an important parameter in flash-cooling macromolecular crystallization solutions", J. Appl. Crystallogr. **38**, 412–419 (2005).

[38] M. Warkentin, V. Berejnov, N. S. Husseini, and R. E. Thorne, "Hyperquenching for protein cryocrystallography", J. Appl. Crystallogr. **39**, 805–811 (2006).

[39] M. Warkentin, V. Stanislavskaia, K. Hammes, and R. E. Thorne, "Cryocrystallography in capillaries: critical glycerol concentrations and cooling rates", J. Appl. Crystallogr. **41**, 791–797 (2008).

[40] H. D. Mertens and D. I. Svergun, "Structural characterization of proteins and complexes using small-angle X-ray solution scattering", J. Struct. Biol. **172**, 128–141 (2010).

[41] D. I. Svergun and M. H. J. Koch, "Small-angle scattering studies of biological macromolecules in solution", Rep. Prog. Phys. **66**, 1735–1782 (2003).

[42] M. Nakasako, "Structural characteristics in protein hydration investigated by cryogenic X-ray crystal structure analyses", J. Biol. Phys. **28**, 129–137 (2002).

[43] R. Henderson, "Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction", Proc. R. Soc. Lond. B Biol. Sci. **241**, 6–8 (1990).

[44] M. Howells, T. Beetz, H. Chapman, C. Cui, J. Holton, C. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, et al., "An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy", J. Electron Spectrosc. Relat. Phenom. **170**, 4–12 (2009).

[45] J. M. Holton, "A beginner's guide to radiation damage", J. Synchrotron Radiat. **16**, 133–142 (2009).

[46] R. D. Leapman and S. Sun, "Cryo-electron energy loss spectroscopy: observations on vitrified hydrated specimens and radiation damage", Ultramicroscopy **59**, 71–79 (1995).

[47] L. Baker and J. Rubinstein, "Radiation damage in electron cryomicroscopy", Methods Enzymol. **481**, 371 (2010).

[48] A. Meents, S. Gutmann, A. Wagner, and C. Schulze-Briese, "Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures", Proc. Natl. Acad. Sci. U.S.A. **107**, 1094–1099 (2010).

[49] T. J. Macke and D. A. Case, "Modeling unusual nucleic acid structures", in *Molecular modeling of nucleic acids*, edited by N. B. Leontis and J. SantaLucia (American Chemical Society, Washington, DC, 1998), pp. 379–393.

[50] B. Vestergaard and S. Hansen, "Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering", J. Appl. Crystallogr. **39**, 797–804 (2006).

[51] D. I. Svergun, "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing", Biophys. J. **76**, 2879–2886 (1999).

[52] V. Vagenende, M. G. S. Yap, and B. L. Trout, "Mechanisms of protein stabilization and prevention of protein aggregation by glycerol", Biochemistry **48**, 11084–11096 (2009).

[53] S. Nakano, H. T. Karimata, Y. Kitagawa, and N. Sugimoto, "Facilitation of RNA enzyme activity in the molecular crowding media of cosolutes", J. Am. Chem. Soc. **131**, 16881–16888 (2009).

[54] C. H. Spink, N. Garbett, and J. B. Chaires, "Enthalpies of DNA melting in the presence of osmolytes", Biophys. Chem. **126**, 176–185 (2007).

[55] D. Miyoshi, H. Karimata, and N. Sugimoto, "Factors regulating thermodynamic stability of DNA structures under molecular crowding conditions", Nucleic Acids Symp. Ser. **50**, 203–204 (2006).

[56] S. Nakano, L. Wu, H. Oka, H. T. Karimata, T. Kirihata, Y. Sato, S. Fujii, H. Sakai, M. Kuwahara, M. Kuwahara, H. Sawai, and N. Sugimoto, "Conformation and the sodium

ion condensation on DNA and RNA structures in the presence of a neutral cosolute as a mimic of the intracellular media", Mol. BioSyst. **4**, 579–588 (2008).

[57] S. Kriminski, M. Kazmierczak, and R. E. Thorne, "Heat transfer from protein crystals: implications for flash-cooling and X-ray beam heating", Acta Crystallogr. D Biol. Crystallogr. **59**, 697–708 (2003).

[58] J. B. Hopkins, R. Badeau, M. Warkentin, and R. E. Thorne, "Effect of common cryoprotectants on critical warming rates and ice formation in aqueous solutions", Cryobiology **65**, 169–178 (2012).

APPENDIX A

SUPPLEMENTARY FIGURES FOR CHAPTER 4

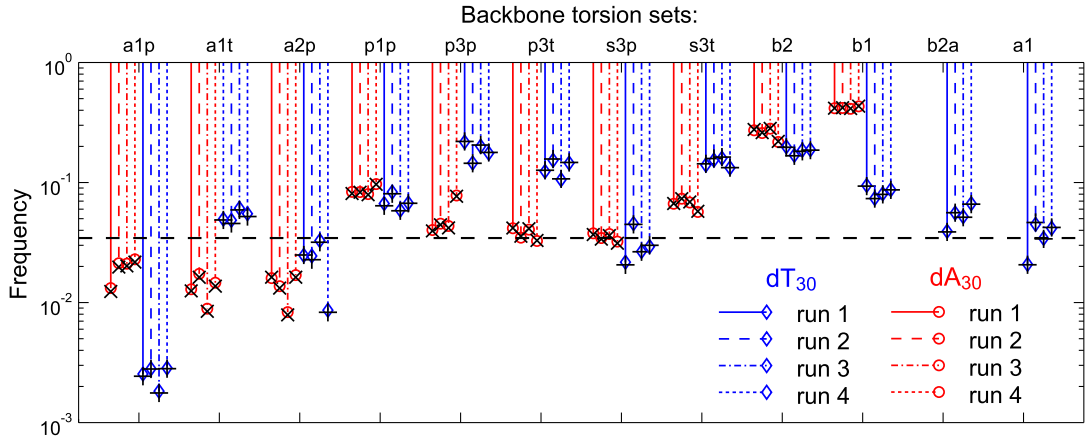

Figure A.1: Torsion angle frequencies in the final pools after 18 rounds of iterative refinement against SAXS data with 20 mM NaCl for dA_{30} (red circles) and dT_{30} (blue diamonds). Ensemble frequencies are also plotted for dT_{30} ($\times$) and dA_{30} ($+$). Four independent runs of torsion angle refinement starting from uniform torsion angle weights were performed for each data set to investigate the uniqueness of the solution. Run 1 corresponds to the results in Figure 4.5.

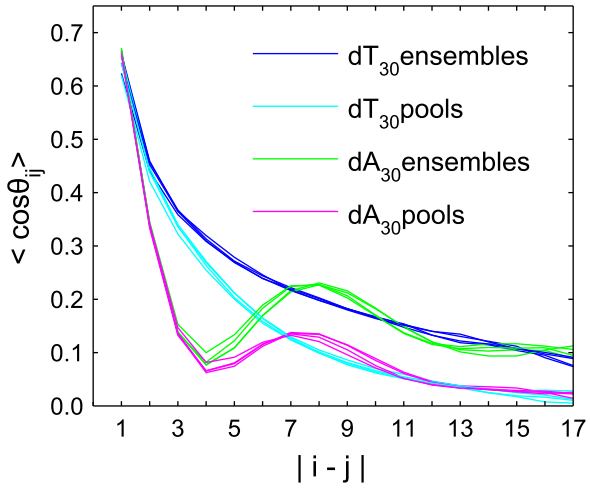

Figure A.2: Orientation correlation functions of dT_{30} and dA_{30} in 20 mM NaCl were derived from the optimized pools and selected ensembles. The results for four independent runs of torsion angle refinement are shown. The selected ensembles have enhanced directional persistence relative to the pools, likely a result of long-ranged electrostatic interactions between charged phosphate groups (the chain generating algorithm includes only steric interactions).

APPENDIX B

SUPPLEMENTARY FIGURES FOR CHAPTER 5

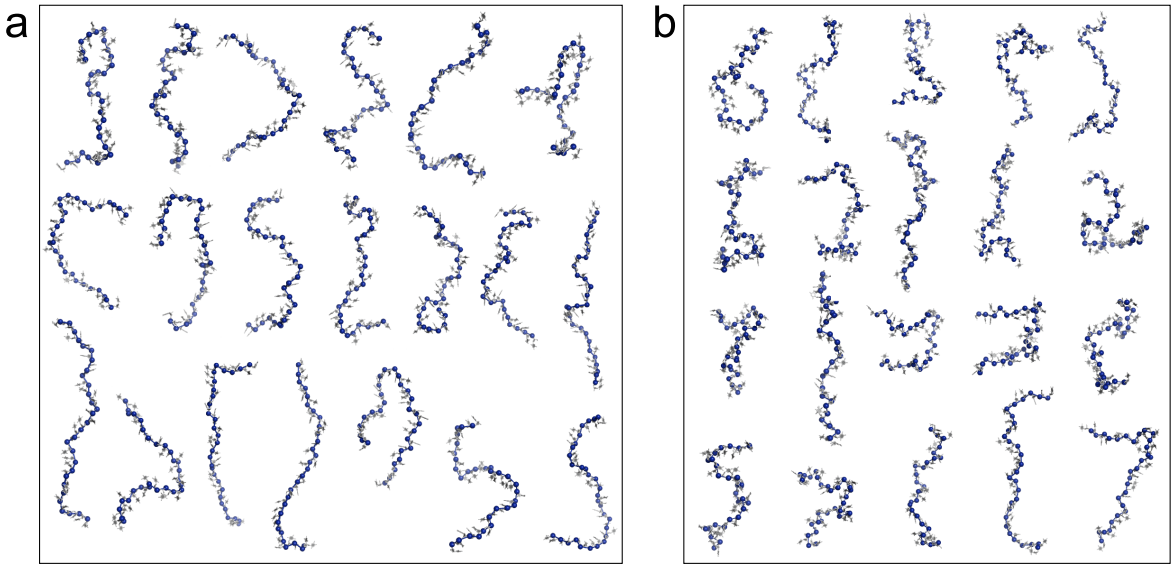

Figure B.1: Ensemble for dT30 in 0 mM Mg (a) and 10 mM Mg (b)

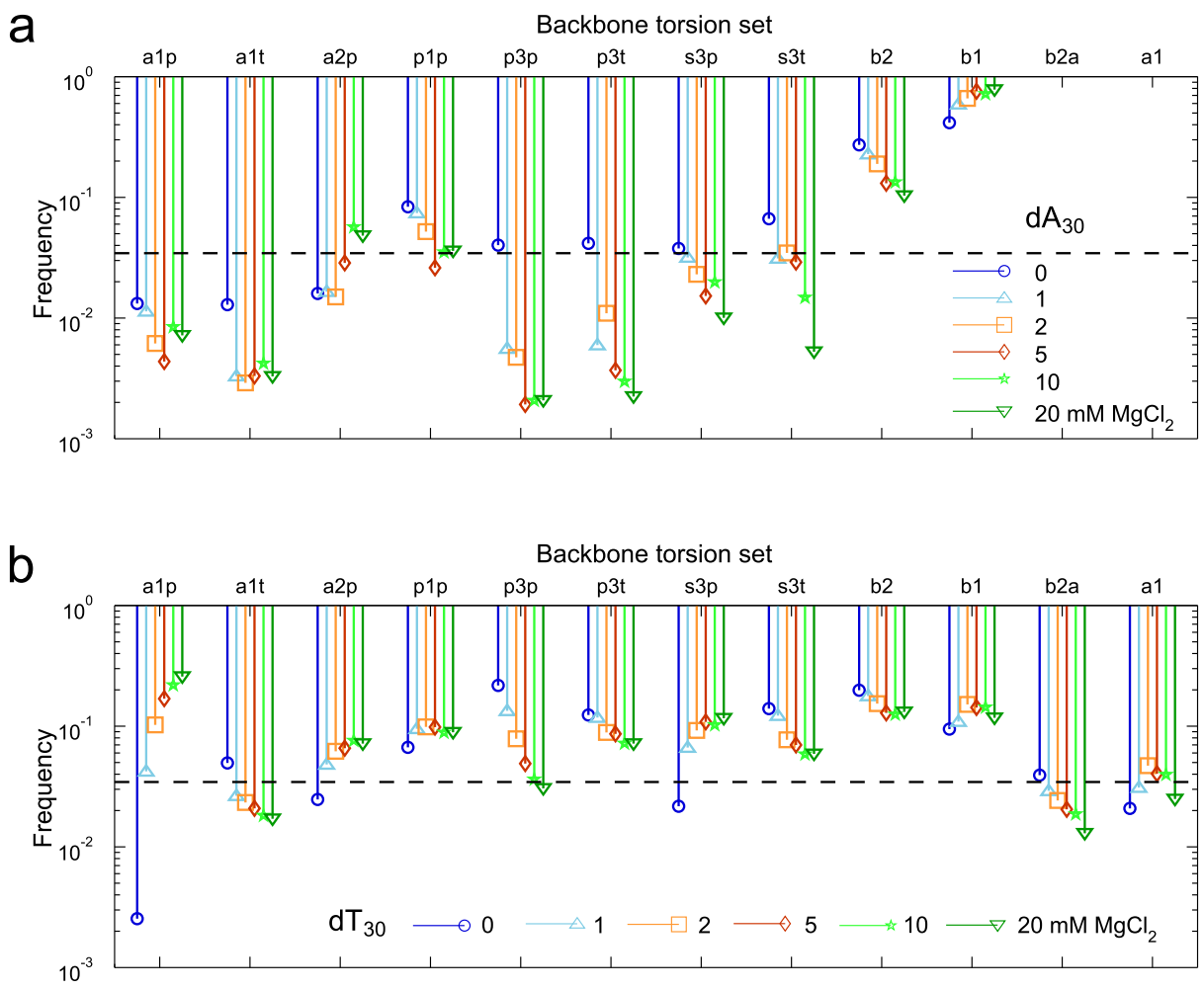

Figure B.2: Backbone torsion angle frequencies for dA_{30} (a) and dT_{30} (b) in 20 mM NaCl with $MgCl_2$ concentrations indicated in the legend. The torsion angle sets are defined in Table 4.1. The horizontal dashed lines indicate a frequency of one per chain.

APPENDIX C

SUPPLEMENTARY FIGURES FOR CHAPTER 6

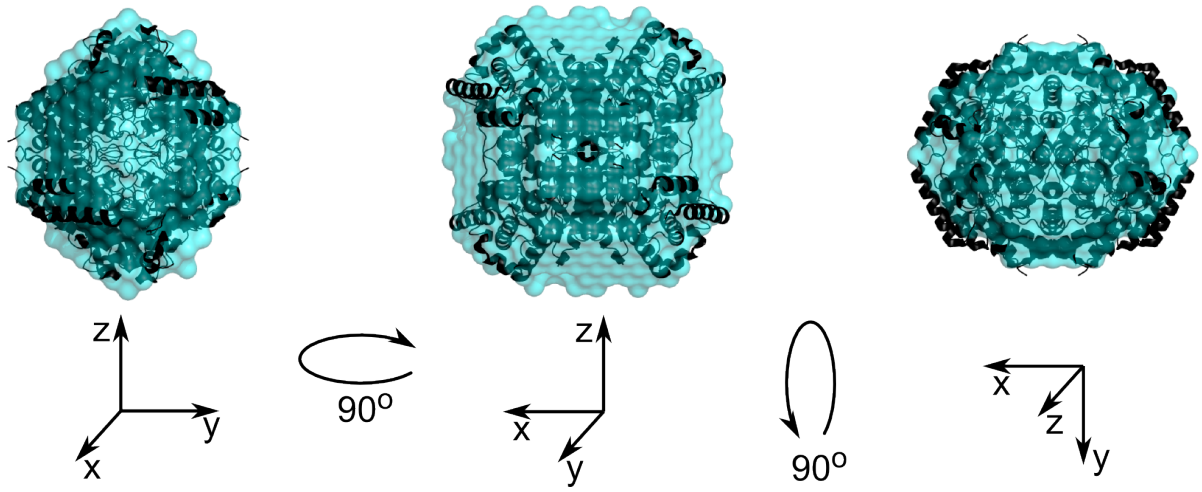

Figure C.1: Cryo-SAXS reconstructions of glucose isomerase with symmetry constraints. Reconstructions of the low resolution envelope were performed using the cryo-SAXS curve of glucose isomerase held in the 1 μL cell over the q-range $0.01 < q < 0.2$ Å^{-1}, enforcing p42 symmetry and oblate anisometry. Ten DAMMIF reconstructions had a mean NSD of 0.531. The filtered average is shown along with the crystal structure PDB 1XIB.

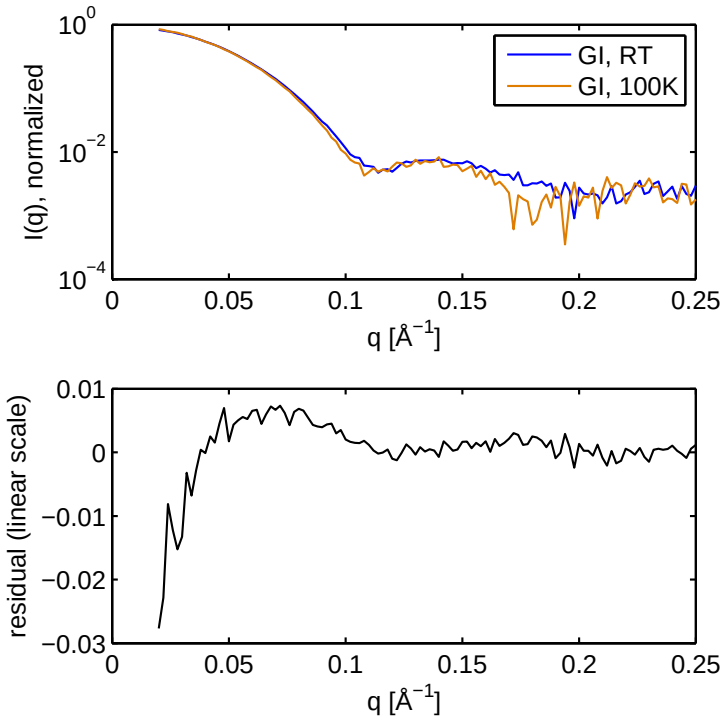

Figure C.2: Direct comparison of glucose isomerase measurements at room temperature (RT, capillary flow cell) and 100 K (cryo-SAXS method, 1 μL cell). The protein concentration was 2 mg/mL, and both samples contained 45% (w/w) PEG 200.

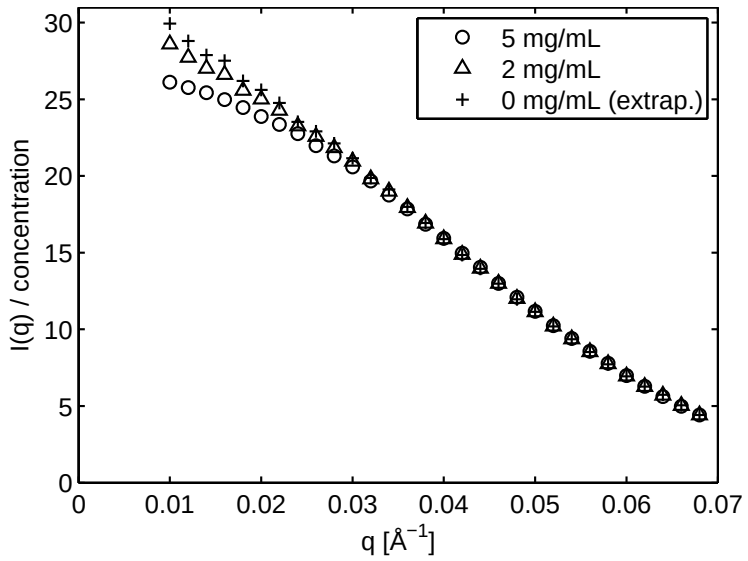

Figure C.3: Observation of interparticle interference in cryocooled glucose isomerase solutions. Cryo-SAXS curves for glucose isomerase solutions at 2 mg/mL and 5 mg/mL concentration were acquired using the 1 μL cell. When the curves are scaled by concentration (matched at mid-q), the 5 mg/mL curve appears below the 2 mg/mL curve at low q, characteristic of repulsive interactions in solution. When extrapolated to zero concentration, the radius of gyration is 34.5 Å.

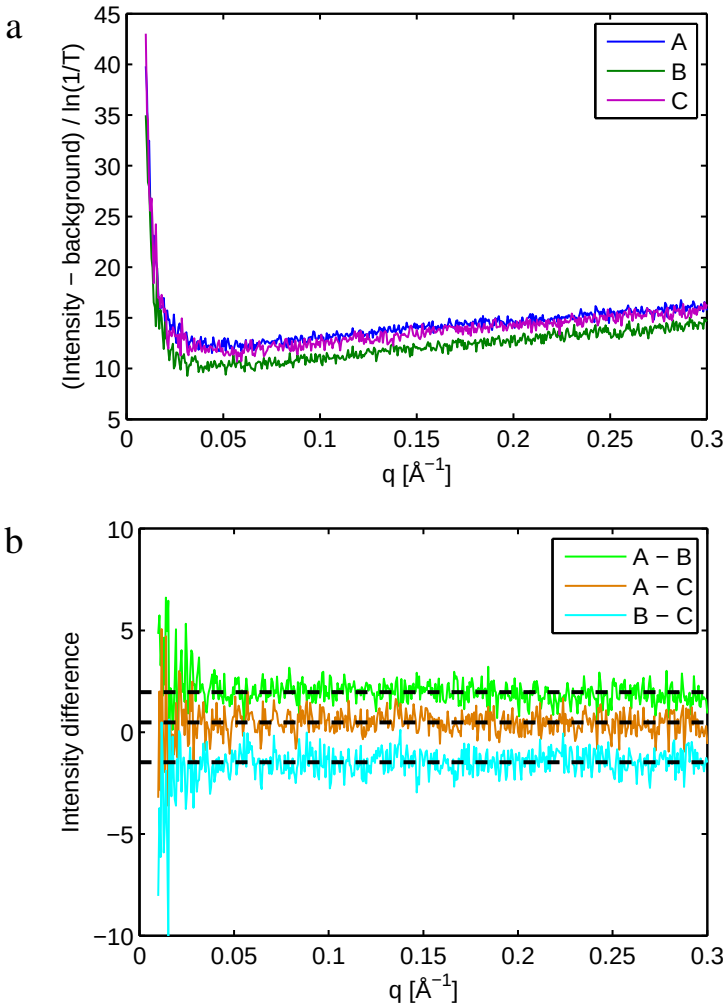

Figure C.4: SAXS patterns of cryo-cooled GI buffer with 45% PEG 200 were acquired from lenticular sub-μl droplets held in a 600μm crystallography loop. (a) Instrumental background scattering was subtracted from each buffer and the curves were normalized by the droplet thickness, according to the background subtraction protocol described in Materials and Methods. These drops do not have the same scattering, indicating that the background subtraction protocol does not adequately compensate for differences in sample geometry. (b) The difference between any pair of curves is well-approximated by a constant offset.

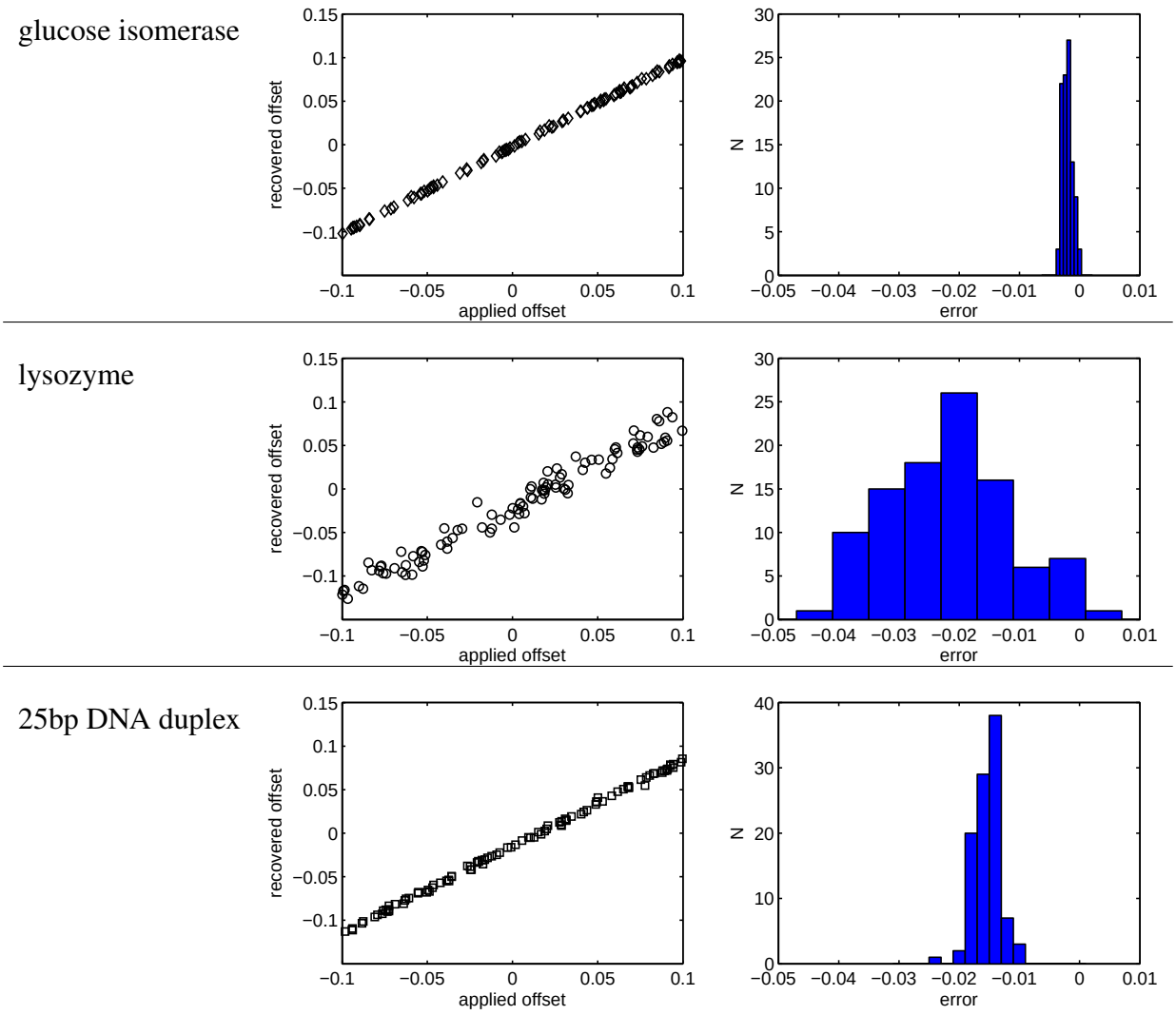

Figure C.5: Performance of the BIFT method for offset correction under simulated experimental conditions. CRYSOL-generated SAXS curves for lysozyme, glucose isomerase (GI), and 24bp DNA were normalized so that $I(0) = 1$, re-sampled at the experimental values of q (0.02 to 0.3 in increments of 0.001 Å^{-1}), and Gaussian random noise corresponding to the experimental error was added. A constant offset was randomly chosen up in the interval [-0.1, 0.1]. For each molecule, 100 such test curves were generated and BIFT was used to find the offset that maximized the evidence of P(r). The applied and recovered offsets are plotted on the left, and the absolute error (recovered - applied) is shown as a histogram on the right. Despite the high level of noise, the BIFT method is fairly reproducible for all simulated data. Lysozyme has the lowest precision of offset recovery; correspondingly, the data had highest relative noise because of Lysozyme's weak scattering. For GI, the SAXS curve approaches zero over the experimental q-range, making the correct offset easier to identify.